# Efficient GMM estimation of spatial dynamic panel data models with fixed effects

Lung-fei Lee
Ohio State University
lee.1777@osu.edu

Jihai Yu
University of Kentucky
jihai.yu@uky.edu

April 28, 2010

**Abstract**

This paper proposes the GMM estimation of the spatial dynamic panel data model with fixed effects when $n$ is large, and $T$ can be large, but small relative to $n$. By eliminating fixed effectsto begin with, we investigate asymptotic properties of the estimators, where exogenous and predetermined variables are used as instruments. For the spatial dynamic panel data model, as compared with the dynamic panel data model, we have not only more linear moment conditions due to spatial effects, but also quadratic moment conditions. We stack up the data and construct the best linear and quadratic moment conditions. An alternative approach is to use separate moment conditions for each period, which gives rise to many moment estimation. We show that these estimators are pnT consistent, asymptotically normal, and can be relatively efficient. We compare these approaches on their finite sample performance by Monte Carlo.

JEL classification: C13; C23; R15

Keywords: Spatial autoregression, Dynamic panels, Fixed effects, Generalized method of moment, Many moments

# 1 Introduction

Dynamic panel data has been studied extensively in recent decades in the literature. It can not only capture dynamics of economic activities but also enable researchers to control unobservable heterogeneity across units. When the number of cross section units $n$ is large, with fixed effects for units, we encounter the incidental parameter problem in Neyman and Scott (1948). As a result, the maximum likelihood estimator (MLE) of the autoregressive coefficient, which is also known as the within estimator, is biased and inconsistent when $n$ tends to infinity but $T$ remains finite (Nickell, 1981; Hsiao, 1986). By taking time differences to eliminate fixed effects in the dynamic equation, the estimation method of instrumental variables (IV) is popular (see Anderson and Hsiao, 1981; Arellano and Bond, 1991; Arellano and Bover, 1995; Blundell and Bond, 1998; Bun and Kiviet, 2006, etc).

When $T$ is finite, the IVs from all the available lag variables may improve, in principle, the asymptotic efficiency of the estimators. When $T$ is moderate or large, however, the many moment issue with a proliferation of IVs will appear. In the literature on IV and generalized method of moments (GMM) estimation with many moment conditions, e.g., in nonlinear simultaneous equations models or conditional moments restrictions models, many moments decrease the variances of the IV or GMM estimates, but increase their biases (see Bekker, 1994; Donald and Newey, 2001; Chao and Swanson, 2005; Han and Phillips, 2006, etc). In the simple dynamic panel data model with fixed effects, when $T$ is moderately large, but small relative to $n$, Alvarez and Arellano (2003) study the many IV estimation and its asymptotic properties. Okui (2009) investigates how to choose the number of instruments to minimize the mean square error (MSE) by extending Donald and Newey (2001) to dynamic panel data models.

Recently, there is a growing literature on spatial panel data models and dynamic panel data models with spatial correlations. By including spatial effects into panel models or dynamic panel models, one can take into account the cross section dependence from contemporaneous or lagged cross section interactions. Kapoor et al. (2007) extend the method of moments estimation to a spatial panel model with error components. Baltagi et al. (2007) consider the testing of spatial and serial dependence in an extended error components model, where serial correlation on each spatial unit over time and spatial dependence across spatial units are in the disturbances. Su and Yang (2007) study the dynamic panel data with spatial error and random effects. These panel models specify spatial correlations by including spatially correlated disturbances and have emphasized error components. In the fixed effects setting, Korniotis (2008) studies a time-space recursive model, where individual time lag and spatial time lag are present, by the least square dummy regression approach. Yu et al. (2007, 2008) and Yu and Lee (2010) study the quasi maximum likelihood (QML) estimation for, respectively, the spatial cointegration, stable, and unit root spatial dynamic panel data (SDPD) models, where individual time lag, spatial time lag and contemporaneous spatial lag are all included.

For the stable SDPD model with fixed effects, the asymptotics of the QML estimation in Yu et al. (2008) is developed under $T \to \infty$ where $T$ cannot be too small relative to $n$. In empirical applications, we might have data sets where $n$ is large while $T$ is relatively small. This motivates our study of GMM estimation of the SDPD model in order to cover the scenario that both $n$ and $T$ can be large, but $T$ is small relative to $n$. The reason for considering the asymptotic with $T \to \infty$ instead of a finite $T$ is that, in this framework, we have the best IV or best GMM estimation with proper designs of IVs and moment conditions.[1] In the QML approach considered in Yu et al (2008), all the parameters including individual fixed effects are jointly estimated, which apparently gives rise to asymptotic biases. In the present paper, we eliminate individual fixed effects first and then consider the IV and GMM estimation of the resulting equation. Specifically, this paper extends the GMM estimation of dynamic panel data models to SDPD models, where we have more linear moment conditions and additional quadratic moment conditions due to spatial effects.

Compared to dynamic panel data models where serial correlation occurs in the time dimension, the SDPD model has correlation in the time dimension as well as spatial correlation across units. In one approach, we stack up the data and use moment conditions in a systematic setting where the IVs have a fixed column dimension for all the periods. In another, we can use separate moment conditions for each time period, which result in many moments. Those many moments not only come from time lags, but are also designed for spatial lags. We focus on the design of estimation methods that can have some asymptotic efficient properties. Normalized asymptotic distributions of IV estimators in the finite moments approaches are properly centered at the true parameter vector. In the many moment approach, normalized asymptotic distributions of IV estimates might not be properly centered or an IV estimator might not be consistent due to the many IV moments (but not directly due to the fixed effects). In contrast to the asymptotics in Yu et al. (2008) where there are ratio conditions on how $T$ and $n$ go to infinity in order that estimates can be consistent or their normalized asymptotic distributions are properly centered, such ratio conditions may no longer be needed with the proposed finite moments estimation methods in the present paper. In the many IVs estimation method, the ratio condition concerns about the number of IVs or moments relative to the total sample size $nT$, but not directly the ratio of $T$ and $n$. However, if the total number of IVs is essentially a function of $T$, then $n$ and $T$ ratio conditions would appear; but in that case, the ratio condition requires that $T$ shall be small relative to $n$. Thus, the many IVs approach is complementary to the QML approach. In other words, the proposed estimation methods can be applied to some scenarios where the $T$ is small relative to $n$, while the QML method might not be, in theory.

The paper is organized as follows. Section 2 introduces the model and discusses moment conditions.

[1] This might not be possible for a fixed effects model when $T$ is assumed to be finite.

Section 3 derives the consistency and asymptotic distribution of GMM estimators when we use finite moment conditions in a systematic setting. Under the framework of $T$ being large, optimal moment conditions can be designed. Section 4 derives the asymptotic properties of GMM estimators when we use many moment conditions. In both Sections 3 and 4, we discuss the asymptotic efficiency of the proposed estimators. Section 5 extends the analysis to the model with also fixed time effects in addition to individual effects. Monte Carlo results for various estimators are provided in Section 6. Section 7 concludes the paper and summarizes the contributions relative to the GMM estimation of the dynamic panel data model and also the QML estimation of similar SDPD models. Some lemmas and proofs are collected in the Appendices.

# 2 The Model and Moment Conditions

## 2.1 The Model

The model we consider in this paper is the SDPD model

$$Y_{nt} = \lambda_0 W_n Y_{nt} + \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} + V_{nt}, \quad t = 1, 2, ..., T, \tag{1}$$

where $Y_{nt} = (y_{1t}, y_{2t}, ..., y_{nt})'$ and $V_{nt} = (v_{1t}, v_{2t}, ..., v_{nt})'$ are $n \times 1$ column vectors, and $v_{it}$'s are *i.i.d.* across $i$ and $t$ with zero mean and variance $\sigma_0^2$. The $W_n$ is an $n \times n$ spatial weights matrix, which is nonstochastic and generates the dependence of $y_{it}$'s across spatial units. $X_{nt}$ is an $n \times k_x$ matrix of nonstochastic regressors and $\mathbf{c}_{n0}$ is an $n \times 1$ column vector of individual effects. The initial values in $Y_{n0}$ are assumed to be observable.

When $n$ is large, to avoid the incidental parameter problem caused by individual effects, they are eliminated by a data transformation. Let $[F_{T,T-1}, \frac{1}{\sqrt{T}} l_T]$ be the orthonormal matrix of the eigenvectors of $J_T = (I_T - \frac{1}{T} l_T l_T')$, where $F_{T,T-1}$ is the $T \times (T-1)$ eigenvectors matrix corresponding to the eigenvalues of one and $l_T$ is the $T$-dimensional vector of ones. The $n \times T$ matrix of dependent variables $[Y_{n1}, Y_{n2}, \cdots, Y_{nT}]$ can be transformed into the $n \times (T-1)$ matrix $[Y_{n1}^*, Y_{n2}^*, \cdots, Y_{n,T-1}^*] = [Y_{n1}, Y_{n2}, \cdots, Y_{nT}]F_{T,T-1}$; and, also, $[Y_{n0}^{(*,-1)}, Y_{n1}^{(*,-1)}, \cdots, Y_{n,T-2}^{(*,-1)}] = [Y_{n0}, Y_{n1}, \cdots, Y_{n,T-1}]F_{T,T-1}$. It is important to note that $Y_{n,t-1}^{(*,-1)}$ and $Y_{n,t-1}^*$ are not equal. Similarly, define $[X_{n1,k}^*, \cdots, X_{n,T-1,k}^*] = [X_{n1,k}, \cdots, X_{nT,k}]F_{T,T-1}$ where $X_{nt,k}$ is the $k$th column of the $n \times k_x$ matrix $X_{nt}$ and $[V_{n1}^*, \cdots, V_{n,T-1}^*] = [V_{n1}, \cdots, V_{nT}]F_{T,T-1}$. Denote $X_{nt}^* = [X_{nt,1}^*, \cdots, X_{nt,k_x}^*]$. As $l_T' F_{T,T-1} = 0$, it follows $[\mathbf{c}_{n0}, \cdots, \mathbf{c}_{n0}]F_{T,T-1} = 0$ so that individual effects are eliminated by the orthonormal transformation. Thus,[2]

$$Y_{nt}^* = \lambda_0 W_n Y_{nt}^* + (\gamma_0 I_n + \rho_0 W_n) Y_{n,t-1}^{(*,-1)} + X_{nt}^* \beta_0 + V_{nt}^*, \quad t = 1, \cdots, T-1. \tag{2}$$

[2]Because $Y_{n,t-1}^{(*,-1)}$ is not $Y_{n,t-1}^*$, (2) does not form a SDPD process by itself. For this reason, an ML or QML approach for (2) is not be feasible.

As $(V_{n1}^{*\prime}, \cdots, V_{n,T-1}^{*\prime})' = (F_{T,T-1}' \otimes I_n)(V_{n1}', \cdots, V_{nT}')'$, $E(V_{n1}^{*\prime}, \cdots, V_{n,T-1}^{*\prime})'(V_{n1}^{*\prime}, \cdots, V_{n,T-1}^{*\prime}) = \sigma_0^2 I_{n(T-1)}$ because $F_{T,T-1}' F_{T,T-1} = I_{T-1}$. Hence, $v_{it}^*$'s are uncorrelated where $v_{it}^*$ is the $i$th element of $V_{nt}^*$. However, we note that $Y_{n,t-1}^{(*,-1)}$ is correlated with $V_{nt}^*$. For this reason, in order to estimate (2) where individual effects are eliminated, IVs are needed for $Y_{n,t-1}^{(*,-1)}$ and $W_n Y_{n,t-1}^{(*,-1)}$ for each $t$ (and also for $W_n Y_{nt}^*$). For this purpose, a convenient selection of $F_{T,T-1}$ is the Helmert transformation. When the Helmert transformation is used, $V_{nt}^* = (\frac{T-t}{T-t+1})^{\frac{1}{2}}[V_{nt} - \frac{1}{T-t}\sum_{h=t+1}^{T} V_{nh}]$ and $Y_{n,t-1}^{(*,-1)} = (\frac{T-t}{T-t+1})^{\frac{1}{2}}[Y_{n,t-1} - \frac{1}{T-t}\sum_{h=t}^{T-1} Y_{nh}]$ depend on current and future variables, but not on the past ones. Thus, in addition to all strictly exogenous variables $X_{ns}$ for $s = 1, \cdots, T-1$, the time lag variables $Y_{n0}, \cdots, Y_{n,t-1}$ can also be used to construct IVs for $Y_{n,t-1}^{(*,-1)}$ as in the literature of dynamic panel data models (Alvarez and Arellano, 2003, etc). Correspondingly, we may use $W_n X_{ns}$ for $s = 1, \cdots, T-1$ and $W_n Y_{ns}$ for $s = 0, ..., t-1$ as IVs for $W_n Y_{n,t-1}^{(*,-1)}$.

## 2.2 Moment Conditions

For the estimation of (2), effective IVs and moment conditions are needed for $W_n Y_{nt}^*$ in addition to those for $Y_{n,t-1}^{(*,-1)}$ and $W_n Y_{n,t-1}^{(*,-1)}$. To motivate the moment conditions for this spatial aspect of the SDPD model, we briefly review the GMM estimation of the cross section spatial autoregressive (SAR) model in order to highlight the particular feature of quadratic moments. For models with spatial interactions, quadratic moments have an important role in efficient estimation.[3]

For the cross section SAR model $Y_n = \lambda_0 W_n Y_n + X_n \beta_0 + V_n$, the reduced form equation is $Y_n = S_n^{-1}(X_n\beta_0 + V_n)$ where $S_n = I_n - \lambda_0 W_n$ and, hence, $W_n Y_n = G_n X_n \beta_0 + G_n V_n$ where $G_n = W_n S_n^{-1}$. The deterministic part $G_n X_n \beta_0 = E(W_n Y_n | X_n)$ is the best IV for a 2SLS approach (Lee, 2003). However, the stochastic component $G_n V_n$ can also be important. One can find IV functions which are correlated with $G_n V_n$ (and hence $W_n Y_n$) but uncorrelated with $V_n$. Lee (2007) shows that the best moment function for this purpose is $(G_n - \frac{tr(G_n)}{n} I_n)V_n$ when elements in $V_n$ are $i.i.d. N(0, \sigma_0^2)$. Lee (2007) proposes the GMM approach based on the linear and quadratic moment conditions $E((G_n X_n \beta_0)' V_n) = 0$ and $E(V_n'(G_n - \frac{tr(G_n)}{n} I_n)V_n) = 0$. The derived GMM estimator is shown to be asymptotically as efficient as the MLE of the SAR model when the disturbances are normally distributed. When the disturbances are non-normal, best linear and quadratic moments also exist, but the expressions can be complicated (see, Liu et al., 2009). To have consistent (but not necessary efficient) estimates, simpler linear and quadratic moments may be used. Kelejian and Prucha (1998) and Kelejian et al. (2004) suggest the use of IVs such as $X_n$, $W_n X_n$ and $W_n^2 X_n$, etc., which approximate $G_n X_n$ in the estimation. One may also use $V_n' W_n V_n$ and/or $V_n'(W_n^2 - \frac{tr(W_n^2)}{n} I_n)V_n$, which are the leading components in the series expansion of $V_n'(G_n - \frac{trG_n}{n} I_n)V_n$, to form quadratic moments. As shown

[3] The use of quadratic moments is motivated by the the likelihood function of the SAR model under normality disturbances (Lee, 2007), as well as the Moran test statistic (Moran, 1950).

in Lee (2007), with $\mathcal{P}_n$ being the class of $n \times n$ constant matrices with zero traces, any finite number of matrices in $\mathcal{P}_n$ can be used in quadratic moment conditions for consistent estimation.

For the GMM estimation of the SDPD model, proper linear moment and quadratic moment conditions for the time and spatial lags, namely, $Y_{n,t-1}^{(*,-1)}$, $W_n Y_{n,t-1}^{(*,-1)}$ and $W_n Y_{nt}^*$, can be revealed from (2). For the linear moments, we can stack up the data and construct moment conditions in a systematic setting. Denote $\mathbf{Q}_{n,T-1} = (Q'_{n1}, \cdots, Q'_{n,T-1})'$ as the IV matrix for the system, where $Q_{nt}$ has a fixed column dimension greater than or equal to $k_x + 3$ for all $t$, e.g., $Q_{nt}$ could be $[Y_{n,t-1}, W_n Y_{n,t-1}, W_n^2 Y_{n,t-1}, X_{nt}^*, W_n X_{nt}^*]$. Then, by denoting $\mathbf{V}_{n,T-1}^* = (V_{n1}^{*\prime}, \cdots, V_{n,T-1}^{*\prime})'$, the linear empirical moments are $\mathbf{Q}'_{n,T-1}\mathbf{V}_{n,T-1}^*$. In another approach, we may use separate moments for each period, where the number of moments might increase over time. Denote $H_{nt}$ as an IV matrix at $t$ consisting of predetermined variables till $t-1$ and all the exogenous variables. For example, $H_{nt}$ can be $(h_{nt}, W_n h_{nt}, \cdots, W_n^{p_n} h_{nt})$ where $h_{nt} = (Y_{n0}, \cdots, Y_{n,t-1}, X_{n1}, \cdots, X_{nT})$ with the integer power $p_n \geq 1$. Then, the linear empirical moments are $H'_{nt} V_{nt}^*$ for $t = 1, \cdots, T-1$. For the quadratic moments, they are designed for the disturbance of $W_n Y_{nt}^*$ in (2). From (2), it follows that

$$W_n Y_{nt}^* = G_n(\gamma_0 I_n + \rho_0 W_n) Y_{n,t-1}^{(*,-1)} + G_n X_{nt}^* \beta_0 + G_n V_{nt}^*, \quad t = 1, \cdots, T-1.$$

This suggests that, in addition to the linear moment conditions, the quadratic moment condition can be $E\mathbf{V}_{n,T-1}^{*\prime}[I_{T-1} \otimes (G_n - \frac{tr(G_n)}{n} I_n)]\mathbf{V}_{n,T-1}^* = 0$. As these moments involve unknown parameters in $G_n$, initial consistent estimates can be obtained from some simpler moment conditions. These generalize the GMM approach for the estimation of dynamic panel data and cross section SAR models to the SDPD model.

Denote $\theta = (\lambda, \gamma, \rho, \beta')'$ and $V_{nt}^*(\theta) = (I_n - \lambda W_n) Y_{nt}^* - (\gamma I_n + \rho W_n) Y_{n,t-1}^{(*,-1)} - X_{nt}^* \beta$. Thus, $\mathbf{V}_{n,T-1}^*(\theta) = (V_{n1}^{*\prime}(\theta), \cdots, V_{n,T-1}^{*\prime}(\theta))'$. In one approach, we propose the following finite moments in the systematic setting:

$$g_{nT}(\theta) = \left(\mathbf{V}_{n,T-1}^{*\prime}(\theta)\mathbf{P}_{n,T-1,1}\mathbf{V}_{n,T-1}^*(\theta), \cdots, \mathbf{V}_{n,T-1}^{*\prime}(\theta)\mathbf{P}_{n,T-1,m}\mathbf{V}_{n,T-1}^*(\theta), \mathbf{V}_{n,T-1}^{*\prime}(\theta)\mathbf{Q}_{n,T-1}\right)', \qquad (3)$$

and, another approach may use separate linear moments for each period, which allows an increasing number of IVs over time:

$$g_{nT}(\theta) = \left(\mathbf{V}_{n,T-1}^{*\prime}(\theta)\mathbf{P}_{n,T-1,1}\mathbf{V}_{n,T-1}^*(\theta), \cdots, \mathbf{V}_{n,T-1}^{*\prime}(\theta)\mathbf{P}_{n,T-1,m}\mathbf{V}_{n,T-1}^*(\theta), \mathbf{V}_{n,T-1}^{*\prime}(\theta) Diag(H_{n1}, \cdots, H_{n,T-1})\right)', \qquad (4)$$

where $Diag(H_{n1}, \cdots, H_{n,T-1})$ is a block diagonal matrix with diagonal blocks $H_{nt}$'s. Here, each $\mathbf{P}_{n,T-1,l}$ for $l = 1, \cdots, m$ is an $n(T-1)$-dimensional nonstochastic square matrix selected from $\mathcal{P}_{n,T-1}$, where $\mathcal{P}_{n,T-1} = I_{T-1} \otimes \mathcal{P}_n$ with $\mathcal{P}_n$ being a class of $n \times n$ matrices with a zero trace. For analytical tractability, we assume that $P_n$ in $\mathcal{P}_n$ is uniformly bounded in row and column sums in absolute value (for short, UB).[4] These

[4] We say a (sequence of $n \times n$) matrix $P_n$ is uniformly bounded in row and column sums if $\sup_{n \geq 1} \|P_n\|_\infty < \infty$ and $\sup_{n \geq 1} \|P_n\|_1 < \infty$, where $\|P_n\|_\infty \equiv \sup_{1 \leq i \leq n} \sum_{j=1}^n |p_{ij,n}|$ is the row sum norm and $\|P_n\|_1 = \sup_{1 \leq j \leq n} \sum_{i=1}^n |p_{ij,n}|$ is the column sum norm.

settings provide general frameworks in which one may discuss the best designs of $Q_{nt}$, $H_{nt}$ and $\mathbf{P}_{n,T-1,l}$.

For (3), the column dimension of $Q_{nt}$ is fixed and is the same for all $t$. For (4), the column dimension of $H_{nt}$ might be increasing in $t$. The latter approach requires careful analysis due to the many moment issue as $T \to \infty$. Hence, appropriately designed $H_{nt}$ might be needed in order that the derived estimate has desirable asymptotic properties. Denote $S_n(\lambda) = I_n - \lambda W_n$ and $S_n \equiv S_n(\lambda_0)$. From the DGP (1), we have $Y_{nt} = A_n^t Y_{n0} + \sum_{h=0}^{t-1} A_n^h S_n^{-1}(\mathbf{c}_{n0} + X_{n,t-h}\beta_0 + V_{n,t-h})$, $t = 1, 2, \cdots, T$, where $A_n = S_n^{-1}(\gamma_0 I_n + \rho_0 W_n)$. For our analysis of the asymptotic properties of estimators, we make the following assumptions.

**Assumption 1.** $W_n$ is a nonstochastic spatial weights matrix with zero diagonals.

**Assumption 2.** The disturbances $\{v_{it}\}$, $i = 1, 2, ..., n$ and $t = 1, 2, ..., T$, are *i.i.d.* across $i$ and $t$ with zero mean, variance $\sigma_0^2$ and $E|v_{it}|^{4+\eta} < \infty$ for some $\eta > 0$.

**Assumption 3.** $S_n(\lambda)$ is invertible for all $\lambda \in \Lambda$, where the parameter space $\Lambda$ is compact and $\lambda_0$ is in the interior of $\Lambda$.

**Assumption 4.** $W_n$ is UB and $\|\lambda_0 W_n\|_\infty < 1$. Also, $S_n^{-1}(\lambda)$ is UB, uniformly in $\lambda \in \Lambda$.[5]

**Assumption 5.** The elements of $X_{nt}$ and $\mathbf{c}_{n0}$ are nonstochastic and bounded, uniformly in $n$ and $t$. Also, $\lim_{n\to\infty} \frac{1}{n(T-1)} \sum_{t=1}^{T-1} X_{nt}^{*\prime} X_{nt}^*$ exists and is nonsingular.

**Assumption 6.** $Y_{n0} = \sum_{h=0}^{h^*} A_n^h S_n^{-1} (\mathbf{c}_{n0} + X_{n,-h}\beta_0 + V_{n,-h})$, where $h^*$ could be finite or infinite.

**Assumption 7.** $\sum_{h=0}^{\infty} abs(A_n^h)$ is UB where $[abs(A_n)]_{ij} = |A_{n,ij}|$.

**Assumption 8.** $n$ goes to infinity.

The zero diagonal assumption on $W_n$ helps the interpretation of the spatial effect as self-influence shall be excluded in practice. Assumption 2 provides regularity assumptions for $v_{it}$. Assumption 3 guarantees that the model is an equilibrium one. Also, the compactness of the parameter space is a condition for theoretical analysis. When $W_n$ is row normalized, a compact subset of $(-1, 1)$ is often taken as the parameter space. In Assumption 4, when $\|\lambda_0 W_n\|_\infty < 1$, $S_n^{-1}$ can be expanded as an infinite series in terms of $W_n$. In many empirical applications of spatial issues, each of the rows of $W_n$ sums to 1, which ensures that all the weights are between 0 and 1. In that case, with $\|W_n\|_\infty = 1$, $|\lambda_0| < 1$ is assumed. The uniform boundedness assumption in Assumption 4 is originated by Kelejian and Prucha (1998, 2001) and also used in Lee (2004, 2007). That $W_n$ and $S_n^{-1}(\lambda)$ are UB is a condition that limits the spatial correlation to a manageable degree. When exogenous variables $X_{nt}$'s are included in the model, it is convenient to assume that they are uniformly bounded as in Assumption 5, and so is $\mathbf{c}_{n0}$. If $X_{nt}$ and $\mathbf{c}_{n0}$ are allowed to be stochastic and unbounded, appropriate moment conditions can be imposed instead. The remaining part of Assumption 5 points out that the regressors of $X_{nt}^*$ are asymptotically linear independent. Assumption 6 specifies the initial condition

[5] This assumption has effectively imposed limited dependence across units. For example, if $\lambda_{0n} = 1 - 1/n$ under $n \to \infty$, it is a near unit root case for a cross sectional spatial autoregressive model and $S_n^{-1}$ will not be UB (see Lee and Yu, 2007).

so that the process may start from a finite or infinite past. Assumption 7 combines the absolute summability condition and the UB condition of the power series of $A_n$, which is essential for the analysis in this paper, as it limits the dependence over time series and across spatial units.[6] Assumption 8 specifies that we have a large number of spatial units, while the time period $T$ could be either large or small. The particular interest in this paper is for the case that $T$ can be large, but small relative to $n$, as the estimation of such a case has not been explicitly covered in the spatial panel literature.

# 3 Asymptotic Properties of GMME with Finite Moments

## 3.1 Consistency and Asymptotic Distribution of GMME

For the moment conditions in (3), identification requires that $\text{plim}_{n\to\infty}\frac{1}{n(T-1)}g_{nT}(\theta) = 0$ should have the unique solution at $\theta_0$. Denote $\mathcal{I}_{t-1}$ as the information set ($\sigma$-algebra) spanned by $(Y_{n0}, \cdots, Y_{n,t-1})$, conditional on $(X_{n1}, \cdots, X_{nT}, \mathbf{c}_{n0})$. Also, denote $\delta = (\gamma, \rho, \beta')'$, $\mathbf{G}_{n,T-1} = I_{T-1} \otimes G_n$ and $\mathbf{Z}^*_{n,T-1} = (Z^{*\prime}_{n1}, \cdots, Z^{*\prime}_{n,T-1})'$ where $Z^*_{nt} = (Y^{(*,-1)}_{n,t-1}, W_n Y^{(*,-1)}_{n,t-1}, X^*_{nt})$ having $k_z = k_x + 2$ columns. The following assumption specifies the identification via rank conditions for the IV estimation.

**Assumption 9.** The $n \times q$ IV matrix $Q_{nt}$ is predetermined such that $E(Q_{nt}|\mathcal{I}_{t-1}) = Q_{nt}$, its column dimension is fixed for all $n$ and $t$ with its elements $O_p(1)$ uniformly in $n$ and $t$, and $\text{plim}_{n\to\infty}\frac{1}{n(T-1)}\mathbf{Q}'_{n,T-1}\mathbf{Q}_{n,T-1}$ is of full rank $q$. Also, $\text{plim}_{n\to\infty}\frac{1}{n(T-1)}\mathbf{Q}'_{n,T-1}[\mathbf{G}_{n,T-1}\mathbf{Z}^*_{n,T-1}\delta_0, \mathbf{Z}^*_{n,T-1}]$ has the full rank $k_z + 1$.

From (2), because $S_n(\lambda) = S_n + (\lambda_0 - \lambda)W_n$, we can expand $V^*_{nt}(\theta)$ as $V^*_{nt}(\theta) = d^*_{nt}(\theta) + S_n(\lambda)S_n^{-1}V^*_{nt}$, where $V^*_{nt} \equiv V^*_{nt}(\theta_0)$ and $d^*_{nt}(\theta) = (\lambda_0 - \lambda)G_n Z^*_{nt}\delta_0 + Z^*_{nt}(\delta_0 - \delta)$. From the linear moment conditions in (3), as $\text{plim}_{n\to\infty}\frac{1}{n(T-1)}\sum_{t=1}^{T-1} Q'_{nt}S_n(\lambda)S_n^{-1}V^*_{nt} = 0$ uniformly in $\theta \in \Theta$ from Lemma 1 (iv), the unique solution of $\text{plim}_{n\to\infty}\frac{1}{n(T-1)}g_{nT}(\theta) = 0$ at $\theta_0$ requires that the equation $\text{plim}_{n\to\infty}\frac{1}{n(T-1)}\mathbf{Q}'_{n,T-1}[\mathbf{G}_{n,T-1}\mathbf{Z}^*_{n,T-1}\delta_0, \mathbf{Z}^*_{n,T-1}]$ $(\lambda_0 - \lambda, (\delta_0 - \delta)')' = 0$ should have a unique solution $\theta_0$. That $\text{plim}_{n\to\infty}\frac{1}{n(T-1)}\mathbf{Q}'_{n,T-1}[\mathbf{G}_{n,T-1}\mathbf{Z}^*_{n,T-1}\delta_0, \mathbf{Z}^*_{n,T-1}]$ has a full rank $k_z + 1$ is a sufficient condition. Because $\mathbf{Z}^*_{n,T-1}$ consists of time and spatial time lags, this condition will, in general, be satisfied as long as $\delta_0 \neq 0$.

Theorem 1 provides the consistency and asymptotic distributions of GMM estimates. The results are valid with either a finite $T$ or $T \to \infty$. As in Hansen's GMM setting (1982), one considers a linear transformation of the moment conditions, $a_{nT}g_{nT}(\theta)$, where $a_{nT}$ is a matrix with its number of rows greater than or equal to $(k_z + 1)$ and $a_{nT}$ is assumed to converge in probability to a constant full rank matrix $a_0$. For the optimal GMM (OGMM) estimation, we need the variance matrix of the moment conditions. Let $vec_D(\mathbf{P}_{n,T-1,j})$ be the column vector formed by diagonal elements of $\mathbf{P}_{n,T-1,j}$ and $vec(\mathbf{P}_{n,T-1,j})$ the column vector formed by stacking the columns of $\mathbf{P}_{n,T-1,j}$. We denote $\omega_{nm,T} = [vec_D(\mathbf{P}_{n,T-1,1}), \cdots, vec_D(\mathbf{P}_{n,T-1,m})]$, and $\Delta_{mn,T} =$

[6] In this paper, we focus only on the stable dynamic model setting, but not unit root or related issues.

$[vec(\mathbf{P}'_{n,T-1,1}),\cdots,vec(\mathbf{P}'_{n,T-1,m})]'[vec(\mathbf{P}^s_{n,T-1,1}),\cdots,vec(\mathbf{P}^s_{n,T-1,m})]$. From Lemma 2, the variance matrix of the moments can be approximated by

$$\Sigma_{nT}=\sigma_0^4\begin{pmatrix}\frac{1}{n(T-1)}\Delta_{nm,T} & \mathbf{0}_{m\times q}\\ \mathbf{0}_{q\times m} & \frac{1}{\sigma_0^2}\frac{1}{n(T-1)}\mathbf{Q}'_{n,T-1}\mathbf{Q}_{n,T-1}\end{pmatrix}+\frac{1}{n(T-1)}\begin{pmatrix}(\mu_4-3\sigma_0^4)\omega'_{nm,T}\omega_{nm,T} & *\\ \mathbf{0}_{q\times m} & \mathbf{0}_{q\times q}\end{pmatrix}, \tag{5}$$

where $\mu_4$ is the fourth moment of $v_{it}$.[7] When $v_{it}$ is normally distributed, the second component of $\Sigma_{nT}$ will be zero because $\mu_4-3\sigma_0^4=0$. For the optimal GMM, $\Sigma_{nT}^{-1}$ is used as $a'_{nT}a_{nT}$. As is shown in Appendix D.1, denoting $\mathbf{P}^s_{n,T-1,j}=\mathbf{P}_{n,T-1,j}+\mathbf{P}'_{n,T-1,j}$, we have $\frac{1}{n(T-1)}\frac{\partial g_{nT}(\hat{\theta}_{nT})}{\partial\theta'}=D_{nT}+o_p(1)$ where

$$D_{nT}=-\frac{1}{n(T-1)}\begin{pmatrix}\sigma_0^2tr(\mathbf{G}'_{n,T-1}\mathbf{P}^s_{n,T-1,1}) & \cdots & \sigma_0^2tr(\mathbf{G}'_{n,T-1}\mathbf{P}^s_{n,T-1,m}) & (\mathbf{G}_{n,T-1}\mathbf{Z}^*_{n,T-1}\delta_0)'\mathbf{Q}_{n,T-1}\\ \mathbf{0}_{k_z\times 1} & \cdots & \mathbf{0}_{k_z\times 1} & \mathbf{Z}^{*\prime}_{n,T-1}\mathbf{Q}_{n,T-1}\end{pmatrix}'.$$

**Theorem 1** *Under Assumptions 1-9, suppose we use the moment conditions in (3) where the nonstochastic matrices* $\mathbf{P}_{n,T-1,l}$ *for* $l=1,\cdots,m$ *are from* $\mathcal{P}_{n,T-1}$ *and* $a_0\text{plim}_{n\to\infty}\frac{1}{n(T-1)}g_{nT}(\theta)=0$ *has a unique root at* $\theta_0$ *in* $\Theta$*, the GMME* $\hat{\theta}_{nT}$ *derived from* $\min_{\theta\in\Theta}g'_{nT}(\theta)a'_{nT}a_{nT}g_{nT}(\theta)$ *is consistent and*

$$\sqrt{n(T-1)}(\hat{\theta}_{nT}-\theta_0)\xrightarrow{d}N(0,\text{plim}_{n\to\infty}\left(D'_{nT}a'_{nT}a_{nT}D_{nT}\right)^{-1}D'_{nT}a'_{nT}a_{nT}\Sigma_{nT}a'_{nT}a_{nT}D_{nT}\left(D_{nT}a'_{nT}a_{nT}D_{nT}\right)^{-1}).$$

*Also, the optimal GMM estimator (OGMME)* $\hat{\theta}_{o,nT}$ *derived from* $\min_{\theta\in\Theta}g'_{nT}(\theta)\Sigma_{nT}^{-1}g_{nT}(\theta)$ *has*

$$\sqrt{n(T-1)}(\hat{\theta}_{o,nT}-\theta_0)\xrightarrow{d}N(0,\text{plim}_{n\to\infty}(D'_{nT}\Sigma_{nT}^{-1}D_{nT})^{-1}). \tag{6}$$

*Suppose that* $\hat{\Sigma}_{nT}^{-1}-\Sigma_{nT}^{-1}=o_p(1)$*, then the feasible OGMME derived from* $\min_{\theta\in\Theta}g'_{nT}(\theta)\hat{\Sigma}_{nT}^{-1}g_{nT}(\theta)$ *has the same asymptotic distribution as (6).*

The OGMME can be compared with the 2SLSE. The 2SLSE of $\theta_0$ is

$$\begin{aligned}\hat{\theta}_{2sl,nT} &= \left[(\mathbf{W}_{n,T-1}\mathbf{Y}^*_{n,T-1},\mathbf{Z}^*_{n,T-1})'\mathbf{M}_{Q,nT}(\mathbf{W}_{n,nT-1}\mathbf{Y}^*_{n,T-1},\mathbf{Z}^*_{n,T-1})\right]^{-1}\\ &\quad\times\left[(\mathbf{W}_{n,T-1}\mathbf{Y}^*_{n,T-1},\mathbf{Z}^*_{n,T-1})'\mathbf{M}_{Q,nT}\mathbf{Y}^*_{n,T-1}\right],\end{aligned}$$

where $\mathbf{M}_{Q,nT}=\mathbf{Q}_{n,T-1}(\mathbf{Q}'_{n,T-1}\mathbf{Q}_{n,T-1})^{-1}\mathbf{Q}_{n,T-1}$. It is consistent and asymptotically normal with the limiting variance matrix $\sigma_0^2\text{plim}_{n\to\infty}(\frac{1}{n(T-1)}(\mathbf{G}_{n,T-1}\mathbf{Z}^*_{n,T-1}\delta_0,\mathbf{Z}^*_{n,T-1})'\mathbf{M}_{Q,nT}(\mathbf{G}_{n,T-1}\mathbf{Z}^*_{n,T-1}\delta_0,\mathbf{Z}^*_{n,T-1}))^{-1}$. The efficiency of the OGMME $\hat{\theta}_{o,nT}$ relative to the 2SLSE is apparent due to additional quadratic moments.

[7] Here, $\Sigma_{nT}$ is not exactly the variance matrix of the moment conditions. While the elements involving the quadratic moment conditions take the expectation form, the elements involving the linear moment conditions take the regular form without expectations or probability limit (because the $Q_{nt}$'s are functions of predetermined variables). However, it has the same limit as the variance matrix. The reason we use such a $\Sigma_{nT}$ is due to its simplicity. See Lemma 2 on how to get $\Sigma_{nT}$.

### 3.2 The Best Linear and Quadratic Moment Conditions

As the quadratic moment conditions and the linear moment conditions of $\mathbf{V}^*_{n,T-1}$ do not interact with each other (see Lemma 2) even though the third moment of $v_{it}$ is not zero, $\Sigma_{nT}$ in (5) is block diagonal. Thus,

$$D'_{nT}\Sigma^{-1}_{nT}D_{nT} = \frac{1}{\sigma_0^2}\frac{1}{n(T-1)}(\mathbf{G}_{n,T-1}\mathbf{Z}^*_{n,T-1}\delta_0, \mathbf{Z}^*_{n,T-1})'\mathbf{M}_{Q,nT}(\mathbf{G}_{n,T-1}\mathbf{Z}^*_{n,T-1}\delta_0, \mathbf{Z}^*_{n,T-1}) + \frac{1}{n(T-1)}\begin{pmatrix} C_{mn,T}(\frac{\mu_4-3\sigma_0^4}{\sigma_0^4}\omega'_{nm,T}\omega_{nm,T}+\Delta_{mn,T})^{-1}C'_{mn,T} & \mathbf{0}_{1\times k_z} \\ \mathbf{0}_{k_z\times 1} & \mathbf{0}_{k_z\times k_z}\end{pmatrix}, \quad (7)$$

where $C_{mn,T} = [tr(\mathbf{P}^s_{n,T-1,1}\mathbf{G}_{n,T-1}), \ldots, tr(\mathbf{P}^s_{n,T-1,m}\mathbf{G}_{n,T-1})]$. When $V_{nt}$ is normally distributed so that $\mu_4 - 3\sigma_0^4 = 0$, the best quadratic moment matrix is $I_{T-1}\otimes\left(G_n - \frac{trG_n}{n}I_n\right)$ by the Cauchy-Schwarz inequality. Without normality, the best quadratic moment matrix shall be (see Appendix D.2), similar to that in Liu et al. (2009),

$$\mathbf{P}_{n,T-1} = I_{T-1}\otimes\left[\left(G_n - \frac{trG_n}{n}I_n\right) - \frac{\mu_4-3\sigma_0^4}{\mu_4-\sigma_0^4}\left(diag(G_n) - \frac{trG_n}{n}I_n\right)\right], \quad (8)$$

where $diag(A)$ denotes the diagonal matrix formed by diagonal elements of a square matrix $A$. When $V_{nt}$ is normally distributed, it simplifies to $I_{T-1}\otimes\left(G_n - \frac{trG_n}{n}I_n\right)$ as expected.

For linear moments, the best $Q_{nt}$ should be the conditional mean $E(W_nY^*_{nt}, Z^*_{nt}|\mathcal{I}_{t-1})$.[8] While this ideal IV matrix might not be directly available, one may design an approximated sequence for it. For that purpose, define $Y^w_{nt} = Y_{nt} - (I_n - A_n)^{-1}S_n^{-1}\mathbf{c}_{n0}$, $\tilde{X}_{n,tT} = \frac{1}{T-t}S_n^{-1}\sum_{h=t}^{T-1}\Phi_{T-h}X_{nh}$ and $\tilde{V}_{n,tT} = \frac{1}{T-t}S_n^{-1}\sum_{h=t}^{T-1}\Phi_{T-h}V_{nh}$ where $\Phi_j = \sum_{h=0}^{j-1}A_n^h$. From Lemma 5, we have

$$Y^{(*,-1)}_{n,t-1} = E(Y^{(*,-1)}_{n,t-1}|\mathcal{I}_{t-1}) - c_{Tt}\tilde{V}_{n,tT}, \quad (9)$$

where $E(Y^{(*,-1)}_{n,t-1}|\mathcal{I}_{t-1}) = \Psi_tY^w_{n,t-1} - c_{Tt}\tilde{X}_{n,tT}\beta_0$, $\Psi_t = c_{Tt}(I_n - \frac{A_n\Phi_{T-t}}{T-t})$ with $c_{Tt} = (\frac{T-t}{T-t+1})^{\frac{1}{2}}$. However, $E(Y^{(*,-1)}_{n,t-1}|\mathcal{I}_{t-1})$ involves the (unknown) fixed effects $(I_n - A_n)^{-1}S_n^{-1}\mathbf{c}_{n0}$ via $Y^w_{n,t-1}$. With initial consistent estimates of $\theta_0$, one may use whole sample observations over time to construct a consistent estimate of $\mathbf{c}_{n0}$, and, then an estimated $Y^w_{nt}$ and an estimated IV for $E(Y^{(*,-1)}_{n,t-1}|\mathcal{I}_{t-1})$. This IV approach is presented in Appendix C, where a large $T$ is crucial to guarantee that $\mathbf{c}_{n0}$ can be consistently estimated. For a finite $T$, such an IV approach could be inconsistent.

As an alternative, at each $t$, one may infer $\mathbf{c}_{n0}$ from observables up to $t-1$, which has the advantage of constructing a consistent IV estimate even if $T$ is finite. As $Y_{ns} = A_nY_{n,s-1} + S_n^{-1}X_{ns}\beta_0 + S_n^{-1}\mathbf{c}_{n0} +$

[8] That $E(W_nY^*_{nt}, Z^*_{nt}|\mathcal{I}_{t-1})$ is the best IV can be seen from the asymptotic variance component of a GMM estimator due to the IVs in Theorem 2.

$S_n^{-1}V_{ns}$, by taking summations over $s = 1$ to $(t-1)$, we have $S_n^{-1}\mathbf{c}_{n0} = \frac{1}{t-1}\sum_{s=1}^{t-1}(Y_{ns} - A_nY_{n,s-1}) - S_n^{-1}\frac{1}{t-1}\sum_{s=1}^{t-1}X_{ns}\beta_0 - S_n^{-1}\frac{1}{t-1}\sum_{s=1}^{t-1}V_{ns}$. Hence, for $t \geq 2$,

$$Y_{n,t-1}^{(*,-1)} = \mathbb{H}_{nt} + [\Psi_t(I_n - A_n)^{-1}S_n^{-1}\frac{1}{t-1}\sum_{s=1}^{t-1}V_{ns} - c_{Tt}\tilde{V}_{n,tT}], \tag{10}$$

where

$$\begin{aligned}\mathbb{H}_{nt} &= \Psi_t[Y_{n,t-1} - (I_n - A_n)^{-1}\frac{1}{t-1}\sum_{s=1}^{t-1}(Y_{ns} - A_nY_{n,s-1})] \\ &\quad +[\Psi_t(I_n - A_n)^{-1}S_n^{-1}\frac{1}{t-1}\sum_{s=1}^{t-1}X_{ns}\beta_0 - c_{Tt}\tilde{X}_{n,tT}\beta_0]. \end{aligned}\tag{11}$$

The best theoretically IV $E(Y_{n,t-1}^{(*,-1)}|\mathcal{I}_{t-1})$ can be approximated by predetermined variables up to the period $t-1$ and exogenous variables up to the period $T-1$ via $\mathbb{H}_{nt}$. Even though $\Psi_t(I_n - A_n)^{-1}S_n^{-1}\frac{1}{t-1}\sum_{s=1}^{t-1}V_{ns}$ is in $\mathcal{I}_{t-1}$ but cannot be observed, it might be ignored. Indeed, it can be small as long as $t$ is far from the initial period. Thus, the approximation can be accurate for those $t$'s far away from the initial period $t = 0$. Hence, we may use $\mathbb{H}_{nt}$ as a desirable IV for $Y_{n,t-1}^*$. For $t = 1$, $E(Y_{n0}^{(*,-1)}|\mathcal{I}_0) = \Psi_1(Y_{n0} - (I_n - A_n)^{-1}S_n^{-1}\mathbf{c}_{n0}) - c_{T1}\tilde{X}_{n,1T}\beta_0$ and we may simply take $\mathbb{H}_{n1} = \Psi_1Y_{n0} - c_{T1}\tilde{X}_{n,1T}\beta_0$. For these IVs with $t$'s close to the initial period $t = 0$, the approximations yield valid IVs but might not be adequate. However, as $T$ is large, the segment with early observations is short relative to the later segment of observations; asymptotically, these IVs are adequate (see Lemma 6). Therefore, the best IV for $Z_{nt}^*$ may be taken as $\mathbb{K}_{nt} \equiv (\mathbb{H}_{nt}, W_n\mathbb{H}_{nt}, X_{nt}^*)$ and the best one for $W_nY_{nt}^*$ is $G_n\mathbb{K}_{nt}\delta_0$. This suggests that we may use

$$\mathbb{Q}_{nt} = (G_n\mathbb{K}_{nt}\delta_0, \mathbb{K}_{nt}) \tag{12}$$

as an IV matrix for $(W_nY_{nt}^*, Z_{nt}^*)$, and its feasible version is

$$\tilde{\mathbb{Q}}_{nt} = (\tilde{G}_n\tilde{\mathbb{K}}_{nt}\tilde{\delta}, \tilde{\mathbb{K}}_{nt}) \tag{13}$$

where $\tilde{G}_n$, $\tilde{\mathbb{K}}_{nt}$ and $\tilde{\delta}$ are feasible counterparts constructed with an initial consistent estimate of $\theta_0$.

**Assumption 10.** The $\mathbf{\Sigma}_{nT,22} = \frac{1}{n(T-1)}(\mathbf{G}_{n,T-1}\mathbf{Z}_{n,T-1}^*\delta_0, \mathbf{Z}_{n,T-1}^*)'(\mathbf{G}_{n,T-1}\mathbf{Z}_{n,T-1}^*\delta_0, \mathbf{Z}_{n,T-1}^*)$ has its probability limit being nonsingular.

**Theorem 2** *Under Assumptions 1-10, suppose we use the moment conditions in (3) where $Q_{nt}$ takes the special form $\tilde{\mathbb{Q}}_{nt}$ in (13) and $\hat{\mathbf{P}}_{n,T-1}$ is estimated[9] from (8). As $n$ and $T$ tend to infinity, the feasible best GMME (BGMME) $\hat{\theta}_{b,nT}$ derived from $\min_{\theta\in\Theta} g_{nT}'(\theta)\hat{\Sigma}_{nT}^{-1}g_{nT}(\theta)$, where $\hat{\Sigma}_{nT}^{-1} - \Sigma_{nT}^{-1} = o_p(1)$, has $\sqrt{n(T-1)}(\hat{\theta}_{b,nT} - \theta_0) \xrightarrow{d} N\left(0, \mathbf{\Sigma}_b^{-1}\right)$ where*

$$\mathbf{\Sigma}_b = \lim_{n\to\infty}\begin{pmatrix} \frac{1}{n(T-1)}tr[\mathbf{P}_{n,T-1}^s\mathbf{G}_{n,T-1}] & \mathbf{0}_{1\times k_z} \\ \mathbf{0}_{k_z\times 1} & \mathbf{0}_{k_z\times k_z}\end{pmatrix} + \frac{1}{\sigma_0^2}\text{plim}_{n\to\infty}\mathbf{\Sigma}_{nT,22}. \tag{14}$$

[9]The $\mathbf{P}_{n,T-1}$ involves the true parameter $\lambda_0$, $\sigma_0^2$ and $\mu_4$, where $\lambda_0$ can be estimated from Theorem 1 and the moments parameters $\sigma_0^2$ and $\mu_4$ can be consistently estimated with the estimated residuals.

For the QML estimator (QMLE) in Yu et al. (2008), it has $O(1/T)$ bias, which can be eliminated but requires the condition that $\frac{T^3}{n} \to \infty$. From Theorem 2, the BGMME does not have a bias term with such an order. Under normality of $V_{nt}$, the BGMME and QMLE have the same asymptotic variance. However, when $V_{nt}$ is not normally distributed, the BGMME with the best IV and best quadratic moment matrix in (8) can be more efficient than the QMLE, because the quadratic moment in the GMM estimation incorporates kurtosis of the disturbances.

We note that when $T$ is finite, the proposed GMME is still consistent and asymptotic normal. However, its limiting variance matrix is the inverse of

$$\mathbf{\Sigma}_c \equiv \lim_{n\to\infty} \begin{pmatrix} \frac{1}{n(T-1)} tr[\mathbf{P}^s_{n,T-1}\mathbf{G}_{n,T-1}] & \mathbf{0}_{1\times k_z} \\ \mathbf{0}_{k_z\times 1} & \mathbf{0}_{k_z\times k_z} \end{pmatrix} \tag{15}$$
$$+\frac{1}{\sigma_0^2}\text{plim}_{n\to\infty}\frac{1}{n(T-1)}(\mathbf{G}_{n,T-1}\mathbf{Z}^*_{n,T-1}\delta_0,\mathbf{Z}^*_{n,T-1})'\mathbf{M}_{\mathbb{Q},nT}(\mathbf{G}_{n,T-1}\mathbf{Z}^*_{n,T-1}\delta_0,\mathbf{Z}^*_{n,T-1}),$$

where $\mathbf{M}_{\mathbb{Q},nT} = \mathbb{Q}_{n,T-1}(\mathbb{Q}'_{n,T-1}\mathbb{Q}_{n,T-1})^{-1}\mathbb{Q}_{n,T-1}$. Thus, when $T$ is finite, the best GMME does not attain the efficiency specified in (14) due to the presence of $\mathbf{M}_{\mathbb{Q},nT}$. This is so, because the best IV $\mathbb{H}_{nt}$ cannot approximate $E(Y^{(*,-1)}_{n,t-1}|\mathcal{I}_{t-1})$ well when $t$ is small.

# 4 Asymptotic Properties of GMME with Many Moments

If we use the moment condition in (4) where the dimension of $H_{nt}$ might increase with $t$ (and also increase with $p_n$, where $p_n$ is the order of spatial expansion of $G_n$), we have the many moment problem (see Bekker 1994, etc) in terms of asymptotic bias. In this section, we investigate the asymptotic properties of the GMM estimator for this approach.

## 4.1 Consistency, Asymptotic Normality and Efficiency of 2SLSE

For the many moment approach, we can use the IV matrix

$$H_{nt} = (h_{nt}, W_n h_{nt}, \cdots, W_n^{p_n} h_{nt}) \text{ with } h_{nt} = (Y_{n0}, ..., Y_{n,t-1}, X_{n1}, ..., X_{nT}) \tag{16}$$

motivated by (11),[10] where the column dimension of $h_{nt}$ is $p_t = k_x T + t$.[11] The $p_n$ (respectively, $p_t$) needs to increase as $n$ (respectively, $t$ and $T$) increases in order to provide adequate approximation to the best theoretical IV. Therefore, the dimension of $H_{nt}$ is $K_t = (p_n + 1)\cdot p_t$. The choice of many moments might have a trade-off between the bias and variance of the GMM estimate, i.e., the larger number of moments might increase the bias of an IV estimator but decrease its variance. In general, we assume that

[10] There are some technical difficulties in the presence of many IVs which involve estimated parameters in the literature, which is also true for our model. Hence, it is desirable to avoid it by using IVs which do not involve estimated parameters.

[11] From (11), an alternative $h_{nt}$ is $(Y_{n0}, \frac{1}{t-2}\sum_{s=1}^{t-2} Y_{ns}, Y_{n,t-1}, \frac{1}{t-1}\sum_{s=1}^{t-1} X_{ns}, X_{nt}, ..., X_{nT})$, which has a smaller dimension.

**Assumption 11.** Both $T$ and $p_n \to \infty$ as $n \to \infty$.

Assumption 11 specifies that, many moments in $H_{nt}$ come out not only from the spatial power series expansion ($p_n \to \infty$) but also from the inclusion of lagged values ($T \to \infty$). As we use a finite number of quadratic moment conditions in the SDPD model, we pay special attention to the linear moments. The additional quadratic moment conditions will not complicate the asymptotic analysis as the two sets of moments do not interact with each other (see Lemma 2).

The $f_{nt} = E(W_n Y_{nt}^*, Z_{nt}^* | \mathcal{I}_{t-1})$ is the best IV for $(W_n Y_{nt}^*, Z_{nt}^*)$. From (2) and (9), $(W_n Y_{nt}^*, Z_{nt}^*) = f_{nt} + u_{nt}$ where

$$f_{nt} = [G_n((\gamma_0 I_n + \rho_0 W_n) E(Y_{n,t-1}^{(*,-1)} | \mathcal{I}_{t-1}) + X_{nt}^* \beta_0), E(Y_{n,t-1}^{(*,-1)} | \mathcal{I}_{t-1}), W_n E(Y_{n,t-1}^{(*,-1)} | \mathcal{I}_{t-1}), X_{nt}^*], \quad (17)$$

$$u_{nt} = [G_n((\gamma_0 I_n + \rho_0 W_n)\eta_{nt} + V_{nt}^*), \eta_{nt}, W_n \eta_{nt}, \mathbf{0}_{n \times k_x}] \text{ with } \eta_{nt} = -c_{Tt} \tilde{V}_{n,tT}. \quad (18)$$

From (10), for $t \geq 2$, the best IV $f_{nt}$ can be approximated by the variables: $Y_{n,t-1}$, $\frac{1}{t-1}\sum_{s=1}^{t-1} Y_{ns}$, $\frac{1}{t-1}\sum_{s=1}^{t-1} Y_{n,s-1}$, $\frac{1}{t-1}\sum_{s=1}^{t-1} X_{ns}$, the exogenous variables afterwards $(X_{nt}, \cdots, X_{nT})$, plus an error component $(\frac{1}{t-1}\sum_{s=1}^{t-1} V_{ns})$, and their spatial expansions.[12] As the elements in $H_{nt}$ contain spatial series involving $W_n$ and $h_{nt}$, the many moments via (16) come out from both spatial and time dimensions. From Lemma 10, we see that $f_{nt}$ can be well approximated by some linear combination of $H_{nt}$ when $t$ is far from the initial period.

The 2SLS estimate is $\hat{\theta}_{2sl,nT} = [\sum_{t=1}^{T-1} (W_n Y_{nt}^*, Z_{nt}^*)' M_{nt} (W_n Y_{nt}^*, Z_{nt}^*)]^{-1} [\sum_{t=1}^{T-1} (W_n Y_{nt}^*, Z_{nt}^*)' M_{nt} Y_{nt}^*]$, where $M_{nt} = H_{nt}(H_{nt}' H_{nt})^+ H_{nt}'$. Thus,

$$\sqrt{n(T-1)}(\hat{\theta}_{2sl,nT} - \theta_0) = \left[\frac{1}{n(T-1)} \sum_{t=1}^{T-1} (f_{nt} + u_{nt})' M_{nt} (f_{nt} + u_{nt})\right]^{-1} \left[\frac{1}{\sqrt{n(T-1)}} \sum_{t=1}^{T-1} (f_{nt} + u_{nt})' M_{nt} V_{nt}^*\right]. \quad (19)$$

From Lemma 6, $\text{plim}_{n\to\infty} \frac{1}{n(T-1)} \sum_{t=1}^{T-1} f_{nt}' f_{nt} = \text{plim}_{n\to\infty} \mathbf{\Sigma}_{nT,22}$ is the probability limit of the first component in (19). As $u_{nt}$ and $V_{nt}^*$ are correlated, the second component in (19) has a non-zero mean. Denote $b_{1,\lambda} = \frac{1}{\sqrt{n(T-1)}} \sigma_0^2 \sum_{t=1}^{T-1} [tr(G_n M_{nt})]$, $b_{2,\lambda} = -\frac{\sigma_0^2}{\sqrt{n(T-1)}} \sum_{t=1}^{T-1} \frac{1}{T+1-t} tr(M_{nt} C_{nTt}' S_n'^{-1} G_n(\gamma_0 I_n + W_n \rho_0))$, $b_{2,\gamma} = -\frac{\sigma_0^2}{\sqrt{n(T-1)}} \sum_{t=1}^{T-1} \frac{1}{T+1-t} tr(M_{nt} C_{nTt}' S_n'^{-1})$ and $b_{2,\rho} = -\frac{\sigma_0^2}{\sqrt{n(T-1)}} \sum_{t=1}^{T-1} \frac{1}{T+1-t} tr(M_{nt} C_{nTt}' S_n'^{-1} W_n)$ with $C_{nTt} = \frac{1}{T-t} \sum_{h=1}^{T-t} h A_n^{h-1}$.

**Theorem 3** *Under Assumptions 1-8 and 10-11, suppose we use many linear moments in (16). Under* $\frac{\sum_{t=1}^{T-1} K_t}{n(T-1)} \to 0$*, the 2SLS* $\hat{\theta}_{2sl,nT}$ *is consistent and*

$$\sqrt{n(T-1)}(\hat{\theta}_{2sl,nT} - \theta_0) - [\hat{H}]^{-1} \cdot (\varphi_1 + \varphi_2) + O_p\left(\frac{\sum_{t=1}^{T-1} \sqrt{K_t}}{\sqrt{n(T-1)}}\right) \xrightarrow{d} N(0, \sigma_0^2 \text{plim}_{n\to\infty} \mathbf{\Sigma}_{nT,22}^{-1}), \quad (20)$$

[12] For $t = 1$, $f_{n1}$ can be approximated by the spatial expansion of $Y_{n0}$ and $(X_{n1}, ..., X_{n,T-1})$.

*where* $\hat{H} = \frac{1}{n(T-1)} \sum_{t=1}^{T-1} (W_n Y_{nt}^*, Z_{nt}^*)' M_{nt} (W_n Y_{nt}^*, Z_{nt}^*)$, $\varphi_1 = e_1 b_{1,\lambda} = O_p\left(\frac{\sum_{t=1}^{T-1} K_t}{\sqrt{n(T-1)}}\right)$ *with* $e_1$ *being the corresponding first unit vector, and* $\varphi_2 = (b_{2,\lambda}, b_{2,\gamma}, b_{2,\rho}, \mathbf{0}_{1\times k_x})' = O_p\left(\frac{1}{\sqrt{n(T-1)}} \sum_{t=1}^{T-1} \frac{K_t}{(T+1-t)(T-t)}\right)$.

*Consequently,*

*(i) if* $\frac{\sum_{t=1}^{T-1} K_t}{\sqrt{n(T-1)}} \to 0$, *then* $\sqrt{n(T-1)}(\hat{\theta}_{2sl,nT} - \theta_0) \xrightarrow{d} N(0, \sigma_0^2 \text{plim}_{n\to\infty} \mathbf{\Sigma}_{nT,22}^{-1})$*;*

*(ii) if* $\frac{\sum_{t=1}^{T-1} K_t}{\sqrt{n(T-1)}} \to c$ *where* $c$ *is a positive finite constant and* $\frac{\max\{K_t : t=1,\ldots,T-1\}}{\sum_{t=1}^{T-1} K_t} \to 0$ *as* $T \to \infty$, *then* $\sqrt{n(T-1)}(\hat{\theta}_{2sl,nT} - \theta_0) - [\hat{H}]^{-1} \cdot \varphi_1 \xrightarrow{d} N(0, \sigma_0^2 \text{plim}_{n\to\infty} \mathbf{\Sigma}_{nT,22}^{-1})$*;*

*(iii) let* $\hat{\theta}_{2sl,nT}^1 = \hat{\theta}_{2sl,nT} - \frac{1}{\sqrt{n(T-1)}} \hat{H}^{-1} \hat{\varphi}_1$ *be a bias corrected estimate, where* $\hat{\varphi}_1$ *is estimated* $\varphi_1$ *with* $\hat{\theta}_{2sl,nT}$. *Then, under the setting in (i), or (ii) and* $\frac{\sum_{t=1}^{T-1} \sqrt{K_t}}{\sqrt{n(T-1)}} \to 0$, $\sqrt{n(T-1)}(\hat{\theta}_{2sl,nT}^1 - \theta_0) \xrightarrow{d} N(0, \sigma_0^2 \text{plim}_{n\to\infty} \mathbf{\Sigma}_{nT,22}^{-1})$.

From Theorem 3, we see that the 2SLS estimate might not be consistent if we have too many moments such that $\frac{\sum_{t=1}^{T-1} K_t}{n(T-1)}$ is not small. Here, the bias $\varphi_1$ in the asymptotic expansion is caused by the endogeneity of the spatial lag, which is of the order $\frac{\sum_{t=1}^{T-1} K_t}{n(T-1)}$ after being rescaled by $\sqrt{n(T-1)}$. The $\varphi_2$ is caused by the correlation of $Z_{nt}^*$ and $V_{nt}^*$ after the data transformation to eliminate individual effects, and $\frac{\varphi_2}{\sqrt{n(T-1)}}$ is of the order $\frac{K}{n(T-1)}$ where $K = \max\{K_t : t = 1, ..., T-1\}$. Thus, the dominant asymptotic bias of the estimate is caused by the endogeneity of the spatial lag term rather than the dynamic lag term. However, after the bias correction, the dominating bias $\varphi_1$ can be eliminated. Comparing the asymptotic distribution of the bias corrected IV estimate in Theorem 3 with that of the IV component of the finite moments approach in Theorem 2, we see that they have the same asymptotic distribution and, thus, both can asymptotically attain the best IV estimate. The asymptotic efficiency of the many IV estimate, however, requires ratio conditions, in particular, that $\frac{\sum_{t=1}^{T-1} K_t}{\sqrt{n(T-1)}} \to 0$. For this requirement to hold, it is implicit that $T$ has to be small relative to $n$.

## 4.2 Consistency, Asymptotic Distribution and Efficiency of GMME

To increase the efficiency of estimates, quadratic moment conditions can be included as those in Section 3. Thus, the moment conditions are (4) where $H_{nt}$ takes the form in (16). Similar to Section 3, the variance matrix of these moment conditions can be approximated by

$$\Sigma_{nT} = \sigma_0^4 \begin{pmatrix} \frac{1}{n(T-1)} \Delta_{nm,T} & \mathbf{0}_{m\times(\Sigma_{t=1}^{T-1} K_t)} \\ \mathbf{0}_{(\Sigma_{t=1}^{T-1} K_t)\times m} & \frac{1}{\sigma_0^2} \frac{1}{n(T-1)} \mathbf{H}_{n,T-1}' \mathbf{H}_{n,T-1} \end{pmatrix} + \frac{1}{n(T-1)} \begin{pmatrix} (\mu_4 - 3\sigma_0^4)\omega_{nm,T}'\omega_{nm,T} & * \\ \mathbf{0}_{(\Sigma_{t=1}^{T-1} K_t)\times m} & \mathbf{0}_{(\Sigma_{t=1}^{T-1} K_t)\times(\Sigma_{t=1}^{T-1} K_t)} \end{pmatrix} \quad (21)$$

where $\mathbf{H}_{n,T-1} = Diag(H_{n1}, \cdots, H_{n,T-1})$ is the block diagonal matrix with $H_{nt}$ in the $t$th diagonal block. We study the optimal GMM with the objective function $g'_{nT}(\theta)\Sigma_{nT}^{-1}g_{nT}(\theta) = g'_{nT,1}(\theta)\Sigma_{nT,1}^{-1}g_{nT,1}(\theta) + g'_{nT,2}(\theta)\Sigma_{nT,2}^{-1}g_{nT,2}(\theta)$, where $g_{nT}(\theta) = (g'_{nT,1}(\theta), g'_{nT,2}(\theta))'$ so that $g_{nT,1}(\theta)$ is the quadratic moment in (4), $g_{nT,2}(\theta)$ is the linear moment in (4), and $\Sigma_{nT}$ is block diagonal from (21) with $\Sigma_{nT} = Diag(\Sigma_{nT,1}, \Sigma_{nT,2})$.

**Theorem 4** *Under Assumptions 1-8 and 10-11, suppose we use many moment conditions in (4) with $H_{nt}$ in (16) and $\hat{\mathbf{P}}_{n,T-1}$ estimated from (8), the feasible BGMME $\hat{\theta}_{b,nT}$ is consistent under $\frac{\sum_{t=1}^{T-1} K_t}{n(T-1)} \to 0$, and*

$$\sqrt{n(T-1)}(\hat{\theta}_{b,nT} - \theta_0) - [\sigma_0^2 \mathbf{\Sigma}_b]^{-1} \cdot (\varphi_1 + \varphi_2) + O_p\left(\frac{\sum_{t=1}^{T-1}\sqrt{K_t}}{\sqrt{n(T-1)}}\right) \xrightarrow{d} N(0, \mathbf{\Sigma}_b^{-1}),$$

*where $\mathbf{\Sigma}_b$ is in (14).*

*Let $\hat{\theta}^1_{b,nT} = \hat{\theta}_{b,nT} - \frac{1}{\sqrt{n(T-1)}}[\hat{\sigma}^2_{nT}\hat{\mathbf{\Sigma}}_b]^{-1}\hat{\varphi}_1$, under the setting in Theorem 3 (iii), the bias corrected BGMME $\hat{\theta}^1_{b,nT}$ has $\sqrt{n(T-1)}(\hat{\theta}^1_{b,nT} - \theta_0) \xrightarrow{d} N\left(0, \mathbf{\Sigma}_b^{-1}\right)$.*

Thus, the BGMME with many IVs can have the same asymptotic distribution as that in Theorem 2.

# 5 A General Model with Time Dummy Effects

The SDPD model (1) can be generalized to include time dummies:

$$Y_{nt} = \lambda_0 W_n Y_{nt} + \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_t l_n + V_{nt}, \quad t = 1, 2, ..., T, \tag{22}$$

where $\alpha_t$ is a fixed time effect. For estimation, we may first eliminate individual effects by $F_{T,T-1}$, which yields $Y^*_{nt} = \lambda_0 W_n Y^*_{nt} + \gamma_0 Y^*_{n,t-1} + \rho_0 W_n Y^{(*,-1)}_{n,t-1} + X^*_{nt}\beta_0 + \alpha^*_t l_n + V^*_{nt}$, $t = 1, 2, ..., T-1$, where $[\alpha^*_1, \alpha^*_2, \cdots, \alpha^*_{T-1}] = [\alpha_1, \alpha_2, \cdots, \alpha_T]F_{T,T-1}$ can be considered as transformed time effects. We make a further transformation to eliminate those time effects $\alpha^*_t$'s. For that purpose, we shall work on the popular spatial scenario that $W_n$ is row normalized.[13]

**Assumption 1'.** $W_n$ is a row normalized nonstochastic spatial weights matrix with zero diagonals.

Let $J_n = I_n - \frac{1}{n} l_n l'_n$ be the deviation from the group mean over spatial units, and let $[F_{n,n-1}, \frac{1}{\sqrt{n}} l_n]$ be the orthonormal matrix of eigenvectors of $J_n$, where the $n \times (n-1)$ eigenvectors matrix $F_{n,n-1}$ corresponds to the eigenvalues of one and $l_n/\sqrt{n}$ corresponds to the eigenvalue zero. We can transform the $n$-dimensional vector $Y^*_{nt}$ to an $(n-1)$-dimensional vector $Y^{**}_{n-1,t}$ by $Y^{**}_{n-1,t} = F'_{n,n-1}Y^*_{nt}$. With $W_n$ being row normalized, because $F'_{n,n-1}W_n l_n = F'_{n,n-1} l_n = 0$, one has

$$Y^{**}_{n-1,t} = \lambda_0(F'_{n,n-1}W_n F_{n,n-1})Y^{**}_{n-1,t} + \gamma_0 Y^{(**,-1)}_{n-1,t-1} + \rho_0(F'_{n,n-1}W_n F_{n,n-1})Y^{(**,-1)}_{n-1,t-1} + X^{**}_{n-1,t}\beta_0 + V^{**}_{n-1,t}, \tag{23}$$

[13] When $W_n$ is not row normalized, we can still eliminate the transformed time effects; however, we will not have the SAR presentation of (23).

where $X^{**}_{n-1,t,k} = F'_{n,n-1}X^*_{nt,k}$ and $V^{**}_{n-1,t} = F'_{n,n-1}V^*_{nt}$. Because $(V^{**\prime}_{n-1,1}, ..., V^{**\prime}_{n-1,T-1})' = (F'_{T,T-1} \otimes F'_{n,n-1})(V'_{n1}, ..., V'_{nT})'$, we have $E(V^{**\prime}_{n-1,1}, ..., V^{**\prime}_{n-1,T-1})'(V^{**\prime}_{n-1,1}, \cdots, V^{**}_{n-1,T-1}) = \sigma_0^2 I_{(n-1)(T-1)}$. Hence, elements of $V^{**}_{n-1,t}$ are uncorrelated for all $i$ and $t$. From (23), as $(I_{n-1} - \lambda_0 F'_{n,n-1}W_n F_{n,n-1})$ is invertible (see Lee and Yu, 2010), $Y^{**}_{n-1,t}$ can be expressed as a function of $Y^{(**,-1)}_{n-1,t-1}$, $X^{**}_{n-1,t}$ and $V^{**}_{n-1,t}$.

## 5.1 Finite Moments Approach in the Systematic Setting

For the linear moments, we similarly stack up the data and construct moment conditions via the transformed equation (23). An IV matrix can take the form $\mathring{\mathbf{Q}}_{n-1,T-1} = (\mathring{Q}'_{n-1,1}, \cdots, \mathring{Q}'_{n-1,T-1})'$ where $\mathring{Q}_{n-1,t} = F'_{n,n-1}Q_{nt}$ has a fixed column dimension $q$ greater than or equal to $k_x + 3$. Thus, the linear moments are $\mathring{\mathbf{Q}}'_{n-1,T-1}\mathbf{V}^{**}_{n-1,T-1}(\theta) = \mathbf{Q}'_{n,T-1}\mathbf{J}_{n,T-1}\mathbf{V}^*_{n,T-1}(\theta)$ where $\mathbf{J}_{n,T-1} = I_{T-1} \otimes J_n$ because $F_{n,n-1}F'_{n,n-1} = J_n$. For the quadratic moments, let $\mathring{\mathbf{P}}_{n-1,T-1,j} = I_{T-1} \otimes F'_{n,n-1}P_{nj}F_{n,n-1}$ for some nonstochastic $n \times n$ matrix $P_{nj}$ with the property $tr(P_{nj}J_n) = 0$. The moment conditions would be

$$\mathring{g}_{nT}(\theta) = \begin{pmatrix} \mathbf{V}^{**\prime}_{n-1,T-1}(\theta)\mathring{\mathbf{P}}_{n-1,T-1,1}\mathbf{V}^{**}_{n-1,T-1}(\theta) \\ \vdots \\ \mathbf{V}^{**\prime}_{n-1,T-1}(\theta)\mathring{\mathbf{P}}_{n-1,T-1,m}\mathbf{V}^{**}_{n-1,T-1}(\theta) \\ \mathring{\mathbf{Q}}'_{n-1,T-1}\mathbf{V}^{**}_{n-1,T-1}(\theta) \end{pmatrix} = \begin{pmatrix} \mathbf{V}^{*\prime}_{n,T-1}(\theta)\mathbf{J}_{n,T-1}\mathbf{P}_{n,T-1,1}\mathbf{J}_{n,T-1}\mathbf{V}^*_{n,T-1}(\theta) \\ \vdots \\ \mathbf{V}^{*\prime}_{n,T-1}(\theta)\mathbf{J}_{n,T-1}\mathbf{P}_{n,T-1,m}\mathbf{J}_{n,T-1}\mathbf{V}^*_{n,T-1}(\theta) \\ \mathbf{Q}'_{n,T-1}\mathbf{J}_{n,T-1}\mathbf{V}^*_{n,T-1}(\theta) \end{pmatrix}. \quad (24)$$

For identification, denote $S^F_{n-1}(\lambda) = I_{n-1} - \lambda F'_{n,n-1}W_n F_{n,n-1}$. From (23), $V^{**}_{n-1,t}(\theta)$ can be expanded as $V^{**}_{n-1,t}(\theta) = d^{**}_{n-1,t}(\theta) + S^F_{n-1}(\lambda)(S^F_{n-1})^{-1}V^{**}_{n-1,t}$, where $V^{**}_{n-1,t} \equiv V^{**}_{n-1,t}(\theta_0)$ and $d^{**}_{n-1,t}(\theta) = F'_{n,n-1}[(\lambda_0 - \lambda)G_n Z^*_{nt}\delta_0 + Z^*_{nt}(\delta_0 - \delta)]$, because $F_{n,n-1}W_n l_n = 0$, $[S^F_{n-1}]^{-1} = F'_{n,n-1}S_n^{-1}(\lambda)F_{n,n-1}$ and $S^F_{n-1}(\lambda)[S^F_{n-1}]^{-1} = I_{n-1} + (\lambda_0 - \lambda)F'_{n,n-1}G_n F_{n,n-1}$. These suggest the following identification conditions.

**Assumption 5'.** The elements of $X_{nt}$ and $\mathbf{c}_{n0}$ are nonstochastic and bounded, uniformly in $n$ and $t$. Also, $\lim_{n\to\infty} \frac{1}{n(T-1)} \sum_{t=1}^{T-1} X^{*\prime}_{nt} J_n X^*_{nt}$ exists and is nonsingular.

**Assumption 9'.** The $n \times q$ IV matrix $Q_{nt}$ is predetermined such that $E(Q_{nt}|\mathcal{I}_{t-1}) = Q_{nt}$, its column dimension is fixed for all $n$ and $t$ with its elements $O_p(1)$ uniformly in $n$ and $t$, and $\text{plim}_{n\to\infty} \frac{1}{n(T-1)}\mathbf{Q}'_{n,T-1}\mathbf{J}_{n,T-1}\mathbf{Q}_{n,T-1}$ is of full rank $q$. Also, $\text{plim}_{n\to\infty} \frac{1}{n(T-1)}\mathbf{Q}'_{n,T-1}\mathbf{J}_{n,T-1}[\mathbf{Z}^*_{n,T-1}, \mathbf{G}_{n,T-1}\mathbf{Z}^*_{n,T-1}\delta_0]$ has the full rank $k_z + 1$.

Assumption 9' is similar to Assumption 9, with the additional $\mathbf{J}_{n,T-1}$ involved due to the additional transformation $F_{n,n-1}$ to eliminate time effects. By defining $\mathring{D}_{nT} = -\frac{1}{(n-1)(T-1)} \times$
$$\begin{pmatrix} \sigma_0^2 tr(\mathbf{G}'_{n,T-1}\mathbf{J}_{n,T-1}\mathbf{P}^s_{n,T-1,1}) & \cdots & \sigma_0^2 tr(\mathbf{G}'_{n,T-1}\mathbf{J}_{n,T-1}\mathbf{P}^s_{n,T-1,m}) & (\mathbf{G}_{n,T-1}\mathbf{Z}^*_{n,T-1}\delta_0)'\mathbf{J}_{n,T-1}\mathbf{Q}_{n,T-1} \\ \mathbf{0}_{k_z\times 1} & \cdots & \mathbf{0}_{k_z\times 1} & \mathbf{Z}^{*\prime}_{n,T-1}\mathbf{J}_{n,T-1}\mathbf{Q}_{n,T-1} \end{pmatrix}',$$
$\frac{1}{(n-1)(T-1)}\frac{\partial \mathring{g}_{nT}(\hat{\theta}_{nT})}{\partial \theta'} = \mathring{D}_{nT} + o_p(1)$ similar to Section 3. Let

$$\begin{aligned} \Delta_{mn,T} &= [vec(\mathbf{J}_{n,T-1}\mathbf{P}'_{n,T-1,1}\mathbf{J}_{n,T-1}), \cdots, vec(\mathbf{J}_{n,T-1}\mathbf{P}'_{n,T-1,m}\mathbf{J}_{n,T-1})]' \\ &\quad \times [vec(\mathbf{J}_{n,T-1}\mathbf{P}^s_{n,T-1,1}\mathbf{J}_{n,T-1}), \cdots, vec(\mathbf{J}_{n,T-1}\mathbf{P}^s_{n,T-1,m}\mathbf{J}_{n,T-1})], \end{aligned}$$

and $\omega_{nm,T} = [vec_D(\mathbf{J}_{n,T-1}\mathbf{P}_{n,T-1,1}\mathbf{J}_{n,T-1}), \cdots, vec_D(\mathbf{J}_{n,T-1}\mathbf{P}_{n,T-1,m}\mathbf{J}_{n,T-1})]$. The variance matrix of these quadratic and linear moments can be approximated by

$$\begin{aligned} \mathring{\Sigma}_{nT} &= \sigma_0^4 \begin{pmatrix} \frac{1}{(n-1)(T-1)}\Delta_{nm,T} & \mathbf{0}_{m\times q} \\ \mathbf{0}_{q\times m} & \frac{1}{\sigma_0^2}\frac{1}{(n-1)(T-1)}\mathbf{Q}'_{n,T-1}\mathbf{J}_{n,T-1}\mathbf{Q}_{n,T-1} \end{pmatrix} \\ &+ \frac{1}{(n-1)(T-1)} \begin{pmatrix} (\mu_4 - 3\sigma_0^4)\omega'_{nm,T}\omega_{nm,T} & * \\ \mathbf{0}_{q\times m} & \mathbf{0}_{q\times q} \end{pmatrix}. \end{aligned} \tag{25}$$

**Theorem 5** *Under Assumptions 1', 2-4, 5', 6-8, and 9', suppose we use the moment conditions in (24), the OGMME* $\hat{\theta}_{o,nT}$ *derived from* $\min_{\theta\in\Theta} \mathring{g}'_{nT}(\theta)\mathring{\Sigma}_{nT}^{-1}\mathring{g}_{nT}(\theta)$ *has*

$$\sqrt{(n-1)(T-1)}(\hat{\theta}_{o,nT} - \theta_0) \xrightarrow{d} N(0, \text{plim}_{n\to\infty}(\mathring{D}'_{nT}\mathring{\Sigma}_{nT}^{-1}\mathring{D}_{nT})^{-1}). \tag{26}$$

*Suppose that* $\hat{\mathring{\Sigma}}_{nT}^{-1} - \mathring{\Sigma}_{nT}^{-1} = o_p(1)$, *then the feasible OGMME derived from* $\min_{\theta\in\Theta} \mathring{g}'_{nT}(\theta)\hat{\mathring{\Sigma}}_{nT}^{-1}\mathring{g}_{nT}(\theta)$ *has the same asymptotic distribution in (26).*

**Proof.** The moment conditions in (24) in terms of $\mathbf{V}^*_{n,T-1}(\theta)$ have similar structures as that in Section 3, except for the presence of $J_n$, which eliminates the time dummies. We note that $J_n$ is UB and the multiplication of UB matrices results in a UB matrix. Thus, asymptotic analysis is similar to Theorem 1. ■

For the corresponding best GMM estimation, from Appendix E, the best quadratic moment has

$$\mathring{\mathbf{P}}^*_{n-1,T-1} = I_{T-1} \otimes F'_{n,n-1}P^*_n F_{n,n-1}, \tag{27}$$

where $P^*_n = (G_n - \frac{trG_n}{n-1}J_n) + \frac{(1-\alpha_n)^2}{(\frac{n}{n-2}+\frac{\eta_4-3}{2})}\left[diag(J_nG_nJ_n) - \frac{tr(G_nJ_n)}{n}I_n\right]$ with $\alpha_n = -\frac{2}{n-2} + \sqrt{\frac{n}{n-2}}\sqrt{\frac{n}{n-2} + \frac{\eta_4-3}{2}}$ and $P^*_n$ is the best within the class of matrices such that $tr(P_nJ_n) = 0$. When $V_{nt}$ is normally distributed so that $\eta_4 = 3$, it implies $\alpha_n = 1$ and the best quadratic matrix is reduced to $I_{T-1} \otimes (G_n - \frac{tr(G_nJ_n)}{n-1}J_n)$. For the linear moments, at $t$, the best IV is $E(F'_{n,n-1}[W_nY^*_{nt}, Z^*_{nt}]|\mathcal{I}_{t-1})$ and its feasible version is

$$F'_{n,n-1}(\tilde{G}_n\tilde{\mathbb{K}}_{nt}\tilde{\delta}, \tilde{\mathbb{K}}_{nt}). \tag{28}$$

**Assumption 10'.** The $\mathring{\mathbf{\Sigma}}_{22,nT} = \frac{1}{(n-1)(T-1)}(\mathbf{G}_{n,T-1}\mathbf{Z}^*_{n,T-1}\delta_0, \mathbf{Z}^*_{n,T-1})'\mathbf{J}_{n,T-1}(\mathbf{G}_{n,T-1}\mathbf{Z}^*_{n,T-1}\delta_0, \mathbf{Z}^*_{n,T-1})$ has its probability limit being nonsingular.

**Theorem 6** *Under Assumptions 1', 2-4, 5', 6-8 and 9'-10', suppose we use the moment conditions in (24) where* $Q_{n-1,t}$ *takes the special form in (28) and* $\mathring{\mathbf{P}}^*_{n-1,T-1}$ *is estimated from (27). As n and T tend to infinity, the feasible BGMME* $\hat{\theta}_{b,nT}$ *derived from* $\min_{\theta\in\Theta} \mathring{g}'_{nT}(\theta)\hat{\mathring{\Sigma}}_{nT}^{-1}\mathring{g}_{nT}(\theta)$, *where* $\hat{\mathring{\Sigma}}_{nT}^{-1} - \mathring{\Sigma}_{nT}^{-1} = o_p(1)$, *has* $\sqrt{(n-1)(T-1)}(\hat{\theta}_{b,nT} - \theta_0) \xrightarrow{d} N(0, \mathring{\mathbf{\Sigma}}_b^{-1})$ *where*

$$\mathring{\mathbf{\Sigma}}_b = \lim_{n\to\infty} \begin{pmatrix} \frac{1}{(n-1)(T-1)}tr[\mathring{\mathbf{P}}^{*s}_{n-1,T-1}\mathring{\mathbf{G}}_{n,T-1}] & \mathbf{0}_{1\times k_z} \\ \mathbf{0}_{k_z\times 1} & \mathbf{0}_{k_z\times k_z} \end{pmatrix} + \frac{1}{\sigma_0^2}\text{plim}_{n\to\infty}\mathring{\mathbf{\Sigma}}_{22,nT}, \tag{29}$$

*with* $\mathring{\mathbf{G}}_{n,T-1} = I_{T-1} \otimes F'_{n,n-1}G_nF_{n,n-1}$.

**Proof.** Similar to Theorem 2. ■

### 5.2 Many Moment Approach

For the separate moments approach, we can use $\mathring{H}_{n-1,t} = F'_{n,n-1}H_{nt}$ for each period, where $H_{nt}$ can be from (16). The many moment conditions are

$$\mathring{g}_{nT}(\theta) = \begin{pmatrix} \mathbf{V}'^{**}_{n-1,T-1}(\theta)\mathring{\mathbf{P}}_{n-1,T-1,1}\mathbf{V}^{**}_{n-1,T-1}(\theta) \\ \vdots \\ \mathbf{V}'^{**}_{n-1,T-1}(\theta)\mathring{\mathbf{P}}_{n-1,T-1,m}\mathbf{V}^{**}_{n-1,T-1}(\theta) \\ Diag(\mathring{H}_{n-1,1},\cdots,\mathring{H}_{n-1,T-1})'\mathbf{V}^{**}_{n-1,T-1}(\theta) \end{pmatrix} = \begin{pmatrix} \mathbf{V}^{*\prime}_{n,T-1}(\theta)\mathbf{J}_{n,T-1}\mathbf{P}_{n,T-1,1}\mathbf{J}_{n,T-1}\mathbf{V}^{*}_{n,T-1}(\theta) \\ \vdots \\ \mathbf{V}^{*\prime}_{n,T-1}(\theta)\mathbf{J}_{n,T-1}\mathbf{P}_{n,T-1,m}\mathbf{J}_{n,T-1}\mathbf{V}^{*}_{n,T-1}(\theta) \\ Diag(H_{n1},\cdots,H_{n,T-1})'\mathbf{J}_{n,T-1}\mathbf{V}^{*}_{n,T-1}(\theta) \end{pmatrix}, \tag{30}$$

where

$$\mathring{H}_{n-1,t} = F'_{n,n-1}(h_{nt}, W_n h_{nt},\cdots,W_n^{p_n}h_{nt}), \text{ with } h_{nt} = (Y_{n0},\cdots,Y_{n,t-1},X_{n1},\cdots,X_{nT}). \tag{31}$$

**Theorem 7** *Under Assumptions 1', 2-4, 5', 6-8, 10' and 11, suppose we use moment conditions in (30) with* $\mathring{H}_{n-1,t}$ *in (31) and* $\mathring{\mathbf{P}}^{*}_{n-1,T-1}$ *estimated from (27), the feasible BGMME* $\hat{\theta}_{b,nT}$ *is consistent under* $\frac{\sum_{t=1}^{T-1}K_t}{n(T-1)} \to 0$.

*Let* $\hat{\theta}^1_{b,nT} = \hat{\theta}_{b,nT} - \frac{1}{\sqrt{n(T-1)}}\hat{\mathring{\mathbf{\Sigma}}}_b^{-1}\hat{\varphi}_1$ *where* $\varphi_1 = e_1\mathring{b}_{1,\lambda}$ *with* $\mathring{b}_{1,\lambda} = \frac{1}{\sqrt{n(T-1)}}\sigma_0^2\sum_{t=1}^{T-1}[tr(J_nG_nM_{nt})]$. *Under the setting in Theorem 3 (iii), the bias corrected BGMME* $\hat{\theta}^1_{b,nT}$ *has* $\sqrt{(n-1)(T-1)}(\hat{\theta}^1_{b,nT}-\theta_0) \xrightarrow{d} N(0,\mathring{\mathbf{\Sigma}}_b^{-1})$, *where* $\mathring{\mathbf{\Sigma}}_b$ *is in (29).*

**Proof.** Similar to Theorem 4. ■

# 6 Monte Carlo

We run simulations to investigate the performance of 2SLSEs and GMMEs in Sections 3 and 4 under different values of $n$, $T$ and $\gamma_0$. We also compare them with those of the QMLE in Yu et al. (2008). Samples are generated from (1):

$$Y_{nt} = \lambda_0 W_n Y_{nt} + \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} + V_{nt}, \quad t = 1,2,...,T,$$

using $\theta_0^a = (0.2, 0.1, -0.2, 1)$, $\theta_0^b = (0.2, 0.5, -0.2, 1)$ and $\theta_0^c = (0.2, 0.9, -0.2, 1)$ where $\theta_0 = (\lambda_0, \gamma_0, \rho_0, \beta_0')'$. Hence, $\gamma_0$ takes the values from 0.1 to 0.9 and other parameters are held constant. The $X_{nt}$, $\mathbf{c}_{n0}$ and $V_{nt}$ are generated from independent standard normal distributions and the spatial weights matrix $W_n$ is a rook matrix.[14] We use $T = 5$, 10, 20, and $n = 100$. For each set of generated sample observations, we

[14] We use the rook matrix based on an $r$ board (so that $n = r^2$). The rook matrix represents a square tessellation with a connectivity of four for the inner fields on the chessboard and two and three for the corner and border fields, respectively. Most empirically observed regional structures in spatial econometrics are made up of regions with connectivity close to the range of the rook tessellation.

calculate the GMM estimator $\hat{\theta}_{nT}$ and evaluate the bias $\hat{\theta}_{nT} - \theta_0$. We do this 1000 times to get the empirical bias $\frac{1}{1000}\sum_{i=1}^{1000}(\hat{\theta}_{nT} - \theta_0)_i$. With three different values of $\theta_0$ for each $n$ and $T$, finite sample properties of these estimators are summarized in Tables 1-7. For each case, we report the bias (Bias), empirical standard deviation (SD) and root mean square error (RMSE). For cases where there are outliers, some quantiles are reported instead.

Tables 1 and 2 use finite moment conditions in (3). Table 1 is for the 2SLSE and GMME using $[Y_{n,t-1}, W_nY_{n,t-1}, \cdots, W_n^5Y_{n,t-1}, X_{nt}^*, W_nX_{nt}^*]$ as the IV matrix,[15] where GMME uses additionally $I_{T-1} \otimes (W_n - \frac{trW_n}{n}I_n)$ and $I_{T-1} \otimes (W_n^2 - \frac{trW_n^2}{n}I_n)$ for quadratic moments. Table 2 is for the BGMMEs, where either $\tilde{\mathbb{Q}}_{nt}$ in (13) or $\hat{\mathbb{Q}}_{nt}^a$ in (37) are used as the IV matrix in linear moments, and $I_{T-1} \otimes (\tilde{G}_n - \frac{tr\tilde{G}_n}{n}I_n)$ for quadratic moment, where $\tilde{G}_n$ is estimated with initial estimates from the GMME in Table 1. For further investigation of BGMMEs in Table 2 compared with GMME in Table 1, we also provide the quantiles of those estimates in Tables 3 and 4.

Tables 5 and 6 use many moments, where IV matrices are $Y_{n0}, \cdots, Y_{n,t-1}, X_{n1}, \cdots, X_{nT}$ and their first five spatial lags. Table 5 is the 2SLSE with and without bias correction. Table 6 is the BGMME with and without bias correction, where $I_{T-1} \otimes (\tilde{G}_n - \frac{tr\tilde{G}_n}{n}I_n)$ is used for the quadratic moment and $\tilde{G}_n$ is estimated with initial estimates from Table 5. All the GMMEs are optimum ones as inverses of their variance matrices are used for weighting. Also, Table 7 is MLEs with and without bias correction.

From Table 1 for the 2SLSE and GMME, Biases are small for all the estimates. For both 2SLSE and GMME, as $T$ increases, SDs decrease; as $\gamma_0$ increases, Biases slightly increase on average and SDs increase. The GMME of $\lambda_0$ has a smaller SD than that of the 2SLSE of $\lambda_0$ such that SDs can be reduced by less than a half; but for other estimates, the performance of GMME and 2SLSE are similar. From Table 2, BGMMEs have small Biases. When $T$ increases or $\gamma_0$ decreases, SDs will be smaller. The BGMMEs have smaller SDs than those of GMM in Table 1 for items (1)-(5), but larger SDs for the rest. From the quantiles of those estimates in Tables 3 and 4, BGMMEs are less dispersed in the specified 25%-75% quantile range, and so is the 10%-90% range. Those large SDs in the BGMME compared to the GMME in Table 2 are caused by some outliers of estimates.

From Table 5, the 2SLSE with many IVs has some biases for the estimate of $\gamma_0$ when $T$ is small, and have biases for the estimate of $\lambda_0$. When $T$ is larger or $\gamma_0$ is smaller, SDs are smaller while the changes in Biases are ambiguous. After the bias correction, Biases and SDs are smaller for the estimate of $\lambda_0$, but

[15] The $[Y_{n,t-1}, W_nY_{n,t-1}, \cdots, W_n^3Y_{n,t-1}, X_{nt}^*, W_nX_{nt}^*]$ is also a valid IV matrix. However, we find that the SDs of the estimates would be much reduced by adding $W_n^4Y_{n,t-1}$ and $W_n^5Y_{n,t-1}$ as the IVs. For the current DGP with exogenous variables, the SDs are reduced by 10% on average with more IVs; for the DGP without exogenous variables, the SDs are large with IV matrix $[Y_{n,t-1}, W_nY_{n,t-1}, \cdots, W_n^3Y_{n,t-1}]$. Detailed simulation results with IV matrix $[Y_{n,t-1}, W_nY_{n,t-1}, \cdots, W_n^3Y_{n,t-1}]$ are available in the supplement file upon request, but are not presented here due to limited space. Also, counterparts for Tables 1-7 without exogenous variables are available.

Table 1: 2SLS and GMME Using Finite IVs in the Systematic Setting

| | | | | | 2SLSE | | | | GMME | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $n$ | $T$ | $\theta_0$ | | $\lambda$ | $\gamma$ | $\rho$ | $\beta$ | $\lambda$ | $\gamma$ | $\rho$ | $\beta$ |
| (1) | 100 | 5 | $\theta_0^a$ | Bias | 0.0039 | -0.0021 | 0.0013 | -0.0039 | -0.0011 | -0.0043 | 0.0008 | -0.0045 |
| | | | | SD | 0.0978 | 0.0595 | 0.1026 | 0.0525 | 0.0624 | 0.0620 | 0.1127 | 0.0525 |
| | | | | RMSE | 0.0978 | 0.0596 | 0.1026 | 0.0527 | 0.0621 | 0.0622 | 0.1127 | 0.0527 |
| (2) | 100 | 10 | $\theta_0^a$ | Bias | 0.0046 | -0.0015 | -0.0000 | -0.0018 | 0.0009 | -0.0021 | 0.0001 | -0.0016 |
| | | | | SD | 0.0640 | 0.0356 | 0.0624 | 0.0327 | 0.0378 | 0.0355 | 0.0627 | 0.0325 |
| | | | | RMSE | 0.0641 | 0.0357 | 0.0624 | 0.0328 | 0.0378 | 0.0355 | 0.0627 | 0.0326 |
| (3) | 100 | 20 | $\theta_0^a$ | Bias | 0.0007 | 0.0010 | -0.0002 | -0.0001 | -0.0015 | 0.0008 | -0.0028 | 0.0003 |
| | | | | SD | 0.0427 | 0.0228 | 0.0419 | 0.0231 | 0.0284 | 0.0230 | 0.0455 | 0.0230 |
| | | | | RMSE | 0.0427 | 0.0228 | 0.0419 | 0.0231 | 0.0284 | 0.0230 | 0.0456 | 0.0230 |
| (4) | 100 | 5 | $\theta_0^b$ | Bias | 0.0035 | -0.0071 | 0.0000 | -0.0052 | 0.0010 | -0.0130 | 0.0047 | -0.0068 |
| | | | | SD | 0.1073 | 0.0931 | 0.1449 | 0.0588 | 0.0748 | 0.0952 | 0.1442 | 0.0591 |
| | | | | RMSE | 0.1074 | 0.0934 | 0.1449 | 0.0590 | 0.0748 | 0.0961 | 0.1443 | 0.0595 |
| (5) | 100 | 10 | $\theta_0^b$ | Bias | 0.0041 | -0.0027 | -0.0031 | -0.0020 | 0.0008 | -0.0043 | -0.0013 | -0.0020 |
| | | | | SD | 0.0656 | 0.0484 | 0.0842 | 0.0338 | 0.0398 | 0.0479 | 0.0820 | 0.0336 |
| | | | | RMSE | 0.0657 | 0.0484 | 0.0743 | 0.0338 | 0.0398 | 0.0481 | 0.0820 | 0.0336 |
| (6) | 100 | 20 | $\theta_0^b$ | Bias | 0.0006 | 0.0013 | -0.0017 | 0.0000 | 0.0004 | 0.0014 | -0.0018 | 0.0000 |
| | | | | SD | 0.0433 | 0.0277 | 0.0540 | 0.0232 | 0.0280 | 0.0277 | 0.0523 | 0.0230 |
| | | | | RMSE | 0.0433 | 0.0278 | 0.0540 | 0.0232 | 0.0280 | 0.0277 | 0.0523 | 0.0230 |
| (7) | 100 | 5 | $\theta_0^c$ | Bias | 0.0169 | -0.0747 | 0.0211 | -0.0374 | 0.0130 | -0.1021 | 0.0266 | -0.0487 |
| | | | | SD | 0.2046 | 0.2938 | 0.2708 | 0.1435 | 0.1251 | 0.2431 | 0.1956 | 0.1214 |
| | | | | RMSE | 0.2053 | 0.3032 | 0.2716 | 0.1483 | 0.1258 | 0.2637 | 0.1974 | 0.1308 |
| (8) | 100 | 10 | $\theta_0^c$ | Bias | 0.0099 | -0.0003 | 0.0025 | -0.0016 | 0.0073 | -0.0154 | 0.0047 | -0.0070 |
| | | | | SD | 0.1179 | 0.1662 | 0.1779 | 0.0735 | 0.0782 | 0.1341 | 0.1131 | 0.0607 |
| | | | | RMSE | 0.1183 | 0.1662 | 0.1780 | 0.0735 | 0.0785 | 0.1350 | 0.1132 | 0.0611 |
| (9) | 100 | 20 | $\theta_0^c$ | Bias | -0.0007 | 0.0177 | -0.0049 | 0.0051 | 0.0048 | 0.0101 | -0.0072 | 0.0020 |
| | | | | SD | 0.0613 | 0.0876 | 0.1089 | 0.0363 | 0.0679 | 0.0977 | 0.0831 | 0.0422 |
| | | | | RMSE | 0.0613 | 0.0894 | 0.1091 | 0.0367 | 0.0681 | 0.0982 | 0.0835 | 0.0422 |

Note: 1. $\theta_0^a$= $(0.2, 0.1, -0.2, 1)$, $\theta_0^b$= $(0.2, 0.5, -0.2, 1)$ and $\theta_0^c$= $(0.2, 0.9, -0.2, 1)$.
2. The IV matrix is $[\mathbf{Y}_{n,T-1}, \mathbf{W}_{n,T-1}\mathbf{Y}_{n,T-1}, ..., \mathbf{W}^5_{n,T-1}\mathbf{Y}_{n,T-1}, \mathbf{X}^*_{n,T-1}, \mathbf{W}_{n,T-1}\mathbf{X}^*_{n,T-1}]$.
3. The quadratic matrices are $I_{T-1}\otimes(W_n - \frac{trW_n}{n} I_n)$ and $I_{T-1}\otimes(W_n^2 - \frac{trW_n^2}{n} I_n)$.

Table 2: BGMMEs Using Finite IVs and Best Quadratic Moments

| | | | | | BGMME in Theorem 2 | | | | BGMME in Theorem 8 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $n$ | $T$ | $\theta_0$ | | $\lambda$ | $\gamma$ | $\rho$ | $\beta$ | $\lambda$ | $\gamma$ | $\rho$ | $\beta$ |
| (1) | 100 | 5 | $\theta_0^a$ | Bias | -0.0016 | -0.0007 | -0.0004 | -0.0026 | -0.0003 | 0.0001 | 0.0016 | -0.0027 |
| | | | | SD | 0.0572 | 0.0518 | 0.0919 | 0.0515 | 0.0550 | 0.0420 | 0.0757 | 0.0510 |
| | | | | RMSE | 0.0572 | 0.0518 | 0.0919 | 0.0515 | 0.0550 | 0.0420 | 0.0757 | 0.0511 |
| (2) | 100 | 10 | $\theta_0^a$ | Bias | 0.0002 | -0.0013 | -0.0010 | -0.0014 | -0.0001 | -0.0011 | -0.0004 | -0.0013 |
| | | | | SD | 0.0374 | 0.0291 | 0.0535 | 0.0325 | 0.0382 | 0.0252 | 0.0454 | 0.0325 |
| | | | | RMSE | 0.0374 | 0.0291 | 0.0535 | 0.0326 | 0.0382 | 0.0252 | 0.0454 | 0.0325 |
| (3) | 100 | 20 | $\theta_0^a$ | Bias | 0.0000 | 0.0003 | -0.0014 | -0.0001 | 0.0002 | 0.0005 | -0.0018 | -0.0002 |
| | | | | SD | 0.0280 | 0.0187 | 0.0353 | 0.0231 | 0.0260 | 0.0172 | 0.0336 | 0.0231 |
| | | | | RMSE | 0.0280 | 0.0187 | 0.0353 | 0.0231 | 0.0260 | 0.0172 | 0.0337 | 0.0231 |
| (4) | 100 | 5 | $\theta_0^b$ | Bias | -0.0023 | -0.0028 | -0.0025 | -0.0027 | -0.0019 | 0.0007 | -0.0004 | -0.0021 |
| | | | | SD | 0.0757 | 0.0744 | 0.1286 | 0.0554 | 0.0652 | 0.0493 | 0.0827 | 0.0523 |
| | | | | RMSE | 0.0757 | 0.0744 | 0.1286 | 0.0554 | 0.0653 | 0.0493 | 0.0827 | 0.0523 |
| (5) | 100 | 10 | $\theta_0^b$ | Bias | 0.0017 | -0.0033 | -0.0025 | -0.0015 | -0.0004 | -0.0026 | 0.0005 | -0.0015 |
| | | | | SD | 0.0537 | 0.0368 | 0.0674 | 0.0332 | 0.0414 | 0.0256 | 0.0528 | 0.0329 |
| | | | | RMSE | 0.0538 | 0.0369 | 0.0675 | 0.0333 | 0.0414 | 0.0257 | 0.0528 | 0.0329 |
| (6) | 100 | 20 | $\theta_0^b$ | Bias | -0.0010 | -0.0015 | 0.0010 | -0.0003 | -0.0024 | -0.0013 | 0.0009 | 0.0008 |
| | | | | SD | 0.0335 | 0.0218 | 0.0548 | 0.0234 | 0.0294 | 0.0160 | 0.0399 | 0.0238 |
| | | | | RMSE | 0.0335 | 0.0218 | 0.0548 | 0.0234 | 0.0295 | 0.0161 | 0.0399 | 0.0238 |
| (7) | 100 | 5 | $\theta_0^c$ | Bias | -0.0025 | -0.0693 | 0.0204 | -0.0321 | 0.0064 | -0.0018 | 0.0055 | -0.0025 |
| | | | | SD | 0.2166 | 0.3237 | 0.3221 | 0.1563 | 0.1583 | 0.1304 | 0.1415 | 0.0777 |
| | | | | RMSE | 0.2166 | 0.3310 | 0.3228 | 0.1596 | 0.1585 | 0.1304 | 0.1416 | 0.0777 |
| (8) | 100 | 10 | $\theta_0^c$ | Bias | 0.0058 | -0.0090 | -0.0017 | -0.0043 | 0.0024 | 0.0079 | -0.0026 | 0.0024 |
| | | | | SD | 0.0948 | 0.0964 | 0.1362 | 0.0499 | 0.0980 | 0.0751 | 0.0765 | 0.0444 |
| | | | | RMSE | 0.0949 | 0.0968 | 0.1363 | 0.0501 | 0.0980 | 0.0755 | 0.0765 | 0.0445 |
| (9) | 100 | 20 | $\theta_0^c$ | Bias | 0.0102 | -0.0108 | -0.0084 | -0.0046 | 0.0045 | 0.0040 | -0.0048 | 0.0020 |
| | | | | SD | 0.0947 | 0.0865 | 0.1262 | 0.0538 | 0.1288 | 0.0640 | 0.0923 | 0.0531 |
| | | | | RMSE | 0.0953 | 0.0872 | 0.1265 | 0.0540 | 0.1289 | 0.0642 | 0.0925 | 0.0531 |

Note: 1. $\theta_0^a = (0.2, 0.1, -0.2, 1)$, $\theta_0^b = (0.2, 0.5, -0.2, 1)$ and $\theta_0^c = (0.2, 0.9, -0.2, 1)$.

2. For BGMME in Theorem 2, the IV matrix is $\tilde{\mathbf{Q}}_{n,T-1} = (\tilde{Q}'_{n1}, ..., \tilde{Q}'_{n,T-1})'$ with $\tilde{Q}_{nt}$ in (13).

3. For BGMME in Theorem 8, the IV matrix is $\hat{\mathbf{Q}}^a_{n,T-1} = (\hat{Q}^{a\prime}_{n1}, ..., \hat{Q}^{a\prime}_{n,T-1})'$ with $\hat{Q}^a_{nt}$ in (37).

4. The quadratic matrix is an estimated $I_{T-1} \otimes (\tilde{G}_n - \frac{tr\tilde{G}_n}{n} I_n)$.

Table 3: Quantiles of BGMM Using Finite IVs and Best Quadratic Moments

| | | | | | BGMME in Theorem 2 | | | | Results for GMME from Table 1 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $n$ | $T$ | $\theta_0$ | | $\lambda$ | $\gamma$ | $\rho$ | $\beta$ | $\lambda$ | $\gamma$ | $\rho$ | $\beta$ |
| (1) | 100 | 5 | $\theta_0^a$ | Median | 0.1979 | 0.0981 | -0.1999 | 0.9964 | 0.2015 | 0.0966 | -0.1991 | 0.9936 |
| | | | | 10%Q | 0.1243 | 0.0340 | -0.3185 | 0.9297 | 0.1244 | 0.0179 | -0.3292 | 0.9270 |
| | | | | 25%Q | 0.1628 | 0.0630 | -0.2624 | 0.9618 | 0.1610 | 0.0574 | -0.2717 | 0.9597 |
| | | | | 75%Q | 0.2365 | 0.1336 | -0.1411 | 1.0328 | 0.2393 | 0.1327 | -0.1291 | 1.0318 |
| | | | | 90%Q | 0.2711 | 0.1668 | -0.0859 | 1.0641 | 0.2746 | 0.1714 | -0.0623 | 1.0636 |
| (2) | 100 | 10 | $\theta_0^a$ | Median | 0.2001 | 0.0991 | -0.2007 | 1.0008 | 0.2002 | 0.0984 | -0.2008 | 1.0000 |
| | | | | 10%Q | 0.1543 | 0.0613 | -0.2681 | 0.9562 | 0.1543 | 0.0541 | -0.2801 | 0.9554 |
| | | | | 25%Q | 0.1754 | 0.0774 | -0.2370 | 0.9755 | 0.1756 | 0.0746 | -0.2415 | 0.9769 |
| | | | | 75%Q | 0.2260 | 0.1183 | -0.1638 | 1.0210 | 0.2256 | 0.1210 | -0.1589 | 1.0205 |
| | | | | 90%Q | 0.2482 | 0.1352 | -0.1325 | 1.0397 | 0.2489 | 0.1430 | -0.1179 | 1.0391 |
| (3) | 100 | 20 | $\theta_0^a$ | Median | 0.1995 | 0.1000 | -0.2020 | 1.0000 | 0.1987 | 0.1007 | -0.2025 | 1.0004 |
| | | | | 10%Q | 0.1672 | 0.0761 | -0.2454 | 0.9704 | 0.1666 | 0.0710 | -0.2559 | 0.9709 |
| | | | | 25%Q | 0.1812 | 0.0879 | -0.2239 | 0.9839 | 0.1812 | 0.0849 | -0.2308 | 0.9842 |
| | | | | 75%Q | 0.2174 | 0.1132 | -0.1777 | 1.0166 | 0.2170 | 0.1162 | -0.1730 | 1.0168 |
| | | | | 90%Q | 0.2323 | 0.1246 | -0.1579 | 1.0289 | 0.2320 | 0.1308 | -0.1463 | 1.0289 |
| (4) | 100 | 5 | $\theta_0^b$ | Median | 0.2006 | 0.4955 | -0.2034 | 0.9970 | 0.2013 | 0.4915 | -0.1948 | 0.9930 |
| | | | | 10%Q | 0.1118 | 0.4063 | -0.3570 | 0.9258 | 0.1097 | 0.3751 | -0.3773 | 0.9175 |
| | | | | 25%Q | 0.1537 | 0.4495 | -0.2832 | 0.9586 | 0.1572 | 0.4289 | -0.2926 | 0.9543 |
| | | | | 75%Q | 0.2414 | 0.5422 | -0.1272 | 1.0340 | 0.2476 | 0.5434 | -0.1056 | 1.0306 |
| | | | | 90%Q | 0.2835 | 0.5932 | -0.0416 | 1.0707 | 0.2914 | 0.6020 | -0.0035 | 1.0688 |
| (5) | 100 | 10 | $\theta_0^b$ | Median | 0.2000 | 0.4974 | -0.2032 | 1.0014 | 0.2013 | 0.4962 | -0.2039 | 0.9995 |
| | | | | 10%Q | 0.1498 | 0.4497 | -0.2878 | 0.9550 | 0.1502 | 0.4365 | -0.3022 | 0.9528 |
| | | | | 25%Q | 0.1756 | 0.4722 | -0.2445 | 0.9757 | 0.1743 | 0.4671 | -0.2570 | 0.9759 |
| | | | | 75%Q | 0.2257 | 0.5221 | -0.1590 | 1.0211 | 0.2269 | 0.5266 | -0.1472 | 1.0211 |
| | | | | 90%Q | 0.2506 | 0.5426 | -0.1144 | 1.0396 | 0.2529 | 0.5544 | -0.0940 | 1.0411 |
| (6) | 100 | 20 | $\theta_0^b$ | Median | 0.2009 | 0.4996 | -0.2009 | 0.9997 | 0.2012 | 0.5013 | -0.2022 | 1.0002 |
| | | | | 10%Q | 0.1669 | 0.4722 | -0.2510 | 0.9698 | 0.1660 | 0.4665 | -0.2644 | 0.9701 |
| | | | | 25%Q | 0.1826 | 0.4851 | -0.2265 | 0.9839 | 0.1829 | 0.4822 | -0.2350 | 0.9834 |
| | | | | 75%Q | 0.2185 | 0.5120 | -0.1757 | 1.0162 | 0.2194 | 0.5186 | -0.1672 | 1.0168 |
| | | | | 90%Q | 0.2322 | 0.5232 | -0.1539 | 1.0289 | 0.2325 | 0.5369 | -0.1382 | 1.0290 |
| (7) | 100 | 5 | $\theta_0^c$ | Median | 0.2004 | 0.8464 | -0.1915 | 0.9763 | 0.2143 | 0.7963 | -0.1677 | 0.9507 |
| | | | | 10%Q | -0.0238 | 0.5319 | -0.5293 | 0.8185 | 0.0599 | 0.5314 | -0.4317 | 0.8128 |
| | | | | 25%Q | 0.1016 | 0.7026 | -0.3548 | 0.8946 | 0.1448 | 0.6500 | -0.3138 | 0.8806 |
| | | | | 75%Q | 0.2931 | 0.9834 | -0.0123 | 1.0486 | 0.2885 | 0.9362 | -0.0339 | 1.0183 |
| | | | | 90%Q | 0.4176 | 1.1398 | 0.1900 | 1.1256 | 0.3650 | 1.0705 | 0.0761 | 1.0948 |
| (8) | 100 | 10 | $\theta_0^c$ | Median | 0.2021 | 0.8915 | -0.2030 | 0.9982 | 0.2093 | 0.8831 | -0.2016 | 0.9936 |
| | | | | 10%Q | 0.1150 | 0.7879 | -0.3615 | 0.9376 | 0.1185 | 0.7293 | -0.3373 | 0.9216 |
| | | | | 25%Q | 0.1584 | 0.8450 | -0.2766 | 0.9653 | 0.1613 | 0.8081 | -0.2725 | 0.9548 |
| | | | | 75%Q | 0.2499 | 0.9450 | -0.1346 | 1.0279 | 0.2519 | 0.9649 | -0.1234 | 1.0302 |
| | | | | 90%Q | 0.3095 | 0.9930 | -0.0491 | 1.0545 | 0.2979 | 1.0436 | -0.0471 | 1.0611 |
| (9) | 100 | 20 | $\theta_0^c$ | Median | 0.2014 | 0.8966 | -0.2018 | 0.9989 | 0.2005 | 0.9129 | -0.2107 | 1.0034 |
| | | | | 10%Q | 0.1562 | 0.8420 | -0.2881 | 0.9636 | 0.1495 | 0.8178 | -0.3008 | 0.9620 |
| | | | | 25%Q | 0.1781 | 0.8700 | -0.2436 | 0.9808 | 0.1725 | 0.8642 | -0.2532 | 0.9811 |
| | | | | 75%Q | 0.2272 | 0.9210 | -0.1591 | 1.0169 | 0.2274 | 0.9612 | -0.1605 | 1.0252 |
| | | | | 90%Q | 0.2478 | 0.9466 | -0.1171 | 1.0333 | 0.2563 | 1.0046 | -0.1057 | 1.0468 |

Note: 1. $\theta_0^a$= $(0.2, 0.1, -0.2, 1)$, $\theta_0^b$= $(0.2, 0.5, -0.2, 1)$ and $\theta_0^c$= $(0.2, 0.9, -0.2, 1)$.

2. The IV matrix is $\tilde{\mathbf{Q}}_{n,T-1} = (\tilde{Q}'_{n1}, ..., \tilde{Q}'_{n,T-1})'$ with $\tilde{Q}_{nt}$ in (13).

3. The quadratic matrix is an estimated $I_{T-1} \otimes (\tilde{G}_n - \frac{tr\tilde{G}_n}{n} I_n)$.

Table 4: Quantiles of BGMM Using Alternative Finite IVs and Best Quadratic Moments

| | | | | | BGMME in Theorem 8 | | | | Results for GMME from Table 6 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $n$ | $T$ | $\theta_0$ | | $\lambda$ | $\gamma$ | $\rho$ | $\beta$ | $\lambda$ | $\gamma$ | $\rho$ | $\beta$ |
| (1) | 100 | 5 | $\theta_0^a$ | Median | 0.1972 | 0.0992 | -0.1969 | 0.9964 | 0.1982 | 0.0302 | -0.1706 | 0.9827 |
| | | | | 10%Q | 0.1319 | 0.0469 | -0.2950 | 0.9307 | 0.1305 | -0.0226 | -0.2549 | 0.9173 |
| | | | | 25%Q | 0.1640 | 0.0725 | -0.2482 | 0.9620 | 0.1623 | 0.0029 | -0.2171 | 0.9482 |
| | | | | 75%Q | 0.2370 | 0.1268 | -0.1476 | 0.0320 | 0.2359 | 0.0571 | -0.1232 | 1.0185 |
| | | | | 90%Q | 0.2719 | 0.1548 | -0.1046 | 1.0643 | 0.2686 | 0.0819 | -0.0801 | 1.0496 |
| (2) | 100 | 10 | $\theta_0^a$ | Median | 0.2005 | 0.0987 | -0.1993 | 1.0004 | 0.1919 | 0.0410 | -0.1789 | 0.9961 |
| | | | | 10%Q | 0.1537 | 0.0660 | -0.2571 | 0.9551 | 0.1465 | 0.0119 | -0.2339 | 0.9505 |
| | | | | 25%Q | 0.1750 | 0.0821 | -0.2307 | 0.9765 | 0.1665 | 0.0267 | -0.2074 | 0.9724 |
| | | | | 75%Q | 0.2251 | 0.1161 | -0.1699 | 1.0209 | 0.2171 | 0.0583 | -0.1493 | 1.0160 |
| | | | | 90%Q | 0.2477 | 0.1305 | -0.1418 | 1.0398 | 0.2402 | 0.0715 | -0.1242 | 1.0349 |
| (3) | 100 | 20 | $\theta_0^a$ | Median | 0.2000 | 0.1009 | -0.2003 | 1.0000 | 0.1908 | 0.0726 | -0.1912 | 1.0001 |
| | | | | 10%Q | 0.1670 | 0.0789 | 0.2426 | 0.9701 | 0.1580 | 0.0517 | -0.2300 | 0.9706 |
| | | | | 25%Q | 0.1823 | 0.0889 | -0.2229 | 0.9836 | 0.1732 | 0.0607 | -0.2113 | 0.9840 |
| | | | | 75%Q | 0.2182 | 0.1120 | -0.1808 | 1.0161 | 0.2089 | 0.0833 | -0.1712 | 1.0171 |
| | | | | 90%Q | 0.2330 | 0.1224 | -0.1614 | 1.0286 | 0.2239 | 0.0929 | -0.1554 | 1.0292 |
| (4) | 100 | 5 | $\theta_0^b$ | Median | 0.1976 | 0.4976 | -0.2017 | 0.9961 | 0.1994 | 0.3845 | -0.1648 | 0.9640 |
| | | | | 10%Q | 0.1231 | 0.4380 | -0.3005 | 0.9285 | 0.1297 | 0.3230 | -0.2549 | 0.8974 |
| | | | | 25%Q | 0.1590 | 0.4668 | -0.2520 | 0.9622 | 0.1621 | 0.3532 | -0.2105 | 0.9282 |
| | | | | 75%Q | 0.2420 | 0.5341 | -0.1479 | 1.0319 | 0.2376 | 0.4159 | -0.1128 | 0.9956 |
| | | | | 90%Q | 0.2808 | 0.5646 | -0.1006 | 1.0660 | 0.2712 | 0.4395 | -0.0743 | 1.0305 |
| (5) | 100 | 10 | $\theta_0^b$ | Median | 0.1997 | 0.4983 | -0.2003 | 1.0005 | 0.1931 | 0.4140 | -0.1722 | 0.9873 |
| | | | | 10%Q | 0.1534 | 0.4654 | -0.2604 | 0.9548 | 0.1456 | 0.3831 | -0.2280 | 0.9410 |
| | | | | 25%Q | 0.1749 | 0.4810 | -0.2319 | 0.9756 | 0.1677 | 0.3986 | -0.2018 | 0.9641 |
| | | | | 75%Q | 0.2253 | 0.5141 | -0.1687 | 1.0212 | 0.2170 | 0.4294 | -0.1453 | 1.0074 |
| | | | | 90%Q | 0.2488 | 0.5288 | -0.1413 | 1.0401 | 0.2396 | 0.4445 | -0.1167 | 1.0270 |
| (6) | 100 | 20 | $\theta_0^b$ | Median | 0.1995 | 0.4992 | -0.2002 | 1.0000 | 0.1913 | 0.4594 | -0.1861 | 0.9984 |
| | | | | 10%Q | 0.1624 | 0.4788 | -0.2437 | 0.9703 | 0.1551 | 0.4402 | -0.2250 | 0.9692 |
| | | | | 25%Q | 0.1804 | 0.4885 | -0.2226 | 0.9843 | 0.1731 | 0.4495 | -0.2074 | 0.9818 |
| | | | | 75%Q | 0.2181 | 0.5097 | -0.1793 | 1.0167 | 0.2091 | 0.4697 | -0.1660 | 1.0153 |
| | | | | 90%Q | 0.2318 | 0.5181 | -0.1571 | 1.0311 | 0.2242 | 0.4779 | -0.1471 | 1.0282 |
| (7) | 100 | 5 | $\theta_0^c$ | Median | 0.2034 | 0.8785 | -0.1852 | 0.9925 | 0.1943 | 0.7172 | -0.1584 | 0.9159 |
| | | | | 10%Q | 0.0872 | 0.7952 | -0.3342 | 0.9138 | 0.1210 | 0.6534 | -0.2565 | 0.8485 |
| | | | | 25%Q | 0.1436 | 0.8314 | -0.2599 | 0.9477 | 0.1606 | 0.6898 | -0.2134 | 0.8828 |
| | | | | 75%Q | 0.2630 | 0.9355 | -0.1250 | 1.0380 | 0.2357 | 0.7502 | -0.1084 | 0.9506 |
| | | | | 90%Q | 0.3301 | 1.0166 | -0.0641 | 1.0838 | 0.2722 | 0.7783 | -0.0706 | 0.9857 |
| (8) | 100 | 10 | $\theta_0^c$ | Median | 0.2029 | 0.8996 | -0.1977 | 1.0037 | 0.1928 | 0.7716 | -0.1684 | 0.9524 |
| | | | | 10%Q | 0.1328 | 0.8531 | -0.2791 | 0.9523 | 0.1447 | 0.7425 | -0.2236 | 0.9079 |
| | | | | 25%Q | 0.1679 | 0.8749 | -0.2393 | 0.9769 | 0.1672 | 0.7565 | -0.1963 | 0.9306 |
| | | | | 75%Q | 0.2359 | 0.9301 | -0.1628 | 1.0270 | 0.2170 | 0.7859 | -0.1394 | 0.9731 |
| | | | | 90%Q | 0.2730 | 0.9718 | -0.1263 | 1.0508 | 0.2422 | 0.7985 | -0.1112 | 0.9929 |
| (9) | 100 | 20 | $\theta_0^c$ | Median | 0.2004 | 0.9019 | -0.2031 | 1.0014 | 0.1919 | 0.8387 | -0.1817 | 0.9841 |
| | | | | 10%Q | 0.1582 | 0.8802 | -0.2582 | 0.9674 | 0.1581 | 0.8248 | -0.2188 | 0.9544 |
| | | | | 25%Q | 0.1774 | 0.8899 | -0.2283 | 0.9843 | 0.1717 | 0.8317 | -0.2008 | 0.9673 |
| | | | | 75%Q | 0.2208 | 0.9167 | -0.1786 | 1.0200 | 0.2096 | 0.8463 | -0.1642 | 1.0005 |
| | | | | 90%Q | 0.2431 | 0.9380 | -0.1559 | 1.0369 | 0.2252 | 0.8539 | -0.1486 | 1.0154 |

Note: 1. $\theta_0^a = (0.2, 0.1, -0.2, 1)$, $\theta_0^b = (0.2, 0.5, -0.2, 1)$ and $\theta_0^c = (0.2, 0.9, -0.2, 1)$.

2. The IV matrix is $\hat{\mathbf{Q}}^a_{n,T-1} = (\hat{Q}^{a\prime}_{n1}, ..., \hat{Q}^{a\prime}_{n,T-1})'$ with $\hat{Q}^a_{nt}$ in (37).

3. The quadratic matrix is an estimated $I_{T-1} \otimes (\tilde{G}_n - \frac{tr\tilde{G}_n}{n} I_n)$.

Table 5: 2SLS Using Many Moments

| | $n$ | $T$ | $\theta_0$ | | 2SLSE before Bias Correction $\lambda$ | $\gamma$ | $\rho$ | $\beta$ | 2SLSE after Bias Correction $\lambda$ | $\gamma$ | $\rho$ | $\beta$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (1) | 100 | 5 | $\theta_0^a$ | Bias | 0.0903 | -0.0671 | 0.0415 | -0.0221 | -0.0339 | -0.0706 | 0.0208 | -0.0145 |
| | | | | SD | 0.0789 | 0.0402 | 0.0686 | 0.0512 | 0.0481 | 0.0399 | 0.0683 | 0.0507 |
| | | | | RMSE | 0.1199 | 0.0782 | 0.0802 | 0.0558 | 0.0588 | 0.0811 | 0.0714 | 0.0528 |
| (2) | 100 | 10 | $\theta_0^a$ | Bias | 0.0964 | -0.0538 | 0.0277 | -0.0125 | -0.0497 | -0.0598 | 0.0191 | -0.0041 |
| | | | | SD | 0.0526 | 0.0234 | 0.0422 | 0.0328 | 0.0281 | 0.0232 | 0.0426 | 0.0324 |
| | | | | RMSE | 0.1098 | 0.0587 | 0.0505 | 0.0351 | 0.0570 | 0.0641 | 0.0467 | 0.0326 |
| (3) | 100 | 20 | $\theta_0^a$ | Bias | 0.0953 | -0.0233 | 0.0097 | -0.0065 | -0.0472 | -0.0296 | 0.0097 | 0.0014 |
| | | | | SD | 0.0360 | 0.0164 | 0.0295 | 0.0231 | 0.0193 | 0.0163 | 0.0299 | 0.0228 |
| | | | | RMSE | 0.1019 | 0.0284 | 0.0311 | 0.0240 | 0.0510 | 0.0338 | 0.0314 | 0.0229 |
| (4) | 100 | 5 | $\theta_0^b$ | Bias | 0.0940 | -0.1160 | 0.0191 | -0.0427 | -0.0366 | -0.1166 | 0.0359 | -0.0347 |
| | | | | SD | 0.0811 | 0.0450 | 0.0712 | 0.0517 | 0.0482 | 0.0448 | 0.0717 | 0.0512 |
| | | | | RMSE | 0.1241 | 0.1245 | 0.0737 | 0.0670 | 0.0605 | 0.1249 | 0.0802 | 0.0618 |
| (5) | 100 | 10 | $\theta_0^b$ | Bias | 0.0990 | -0.0842 | -0.0059 | -0.0216 | -0.0506 | -0.0871 | 0.0403 | -0.0129 |
| | | | | SD | 0.0534 | 0.0234 | 0.0423 | 0.0331 | 0.0194 | 0.0149 | 0.0286 | 0.0229 |
| | | | | RMSE | 0.1125 | 0.0874 | 0.0427 | 0.0395 | 0.0580 | 0.0901 | 0.0580 | 0.0351 |
| (6) | 100 | 20 | $\theta_0^b$ | Bias | 0.0966 | -0.0379 | -0.0284 | -0.0088 | -0.0478 | -0.0412 | 0.0290 | -0.0008 |
| | | | | SD | 0.0364 | 0.0150 | 0.0306 | 0.0231 | 0.0194 | 0.0149 | 0.0286 | 0.0229 |
| | | | | RMSE | 0.1032 | 0.0407 | 0.0418 | 0.0247 | 0.0516 | 0.0438 | 0.0408 | 0.0229 |
| (7) | 100 | 5 | $\theta_0^c$ | Bias | 0.0939 | -0.1834 | -0.0075 | -0.0888 | -0.0472 | -0.1812 | 0.0488 | -0.0809 |
| | | | | SD | 0.0869 | 0.0479 | 0.0774 | 0.0527 | 0.0480 | 0.0471 | 0.0739 | 0.0522 |
| | | | | RMSE | 0.1279 | 0.1896 | 0.0778 | 0.1033 | 0.0673 | 0.1872 | 0.0886 | 0.0963 |
| (8) | 100 | 10 | $\theta_0^c$ | Bias | 0.1020 | -0.1290 | -0.0390 | -0.0555 | -0.0548 | -0.1287 | 0.0624 | -0.0469 |
| | | | | SD | 0.0551 | 0.0223 | 0.0471 | 0.0333 | 0.0284 | 0.0221 | 0.0395 | 0.0330 |
| | | | | RMSE | 0.1159 | 0.1309 | 0.0612 | 0.0648 | 0.0617 | 0.1306 | 0.0738 | 0.0573 |
| (9) | 100 | 20 | $\theta_0^c$ | Bias | 0.0987 | -0.0608 | -0.0655 | -0.0229 | -0.0492 | -0.0611 | 0.0492 | -0.0148 |
| | | | | SD | 0.0369 | 0.0112 | 0.0340 | 0.0233 | 0.0195 | 0.0111 | 0.0249 | 0.0231 |
| | | | | RMSE | 0.1054 | 0.0618 | 0.0738 | 0.0327 | 0.0529 | 0.0621 | 0.0551 | 0.0274 |

Note: 1. $\theta_0^a = (0.2, 0.1, -0.2, 1)$, $\theta_0^b = (0.2, 0.5, -0.2, 1)$ and $\theta_0^c = (0.2, 0.9, -0.2, 1)$.
2. The IVs are $Y_{n0}, \cdots, Y_{n,t-1}, X_{n1}, \cdots, X_{nT}$ and their first 5 spatial lags for the period $t$.

Table 6: GMM Using Many IVs and Best Quadratic Moment

| | $n$ | $T$ | $\theta_0$ | | GMME before Bias Correction $\lambda$ | $\gamma$ | $\rho$ | $\beta$ | GMME after Bias Correction $\lambda$ | $\gamma$ | $\rho$ | $\beta$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (1) | 100 | 5 | $\theta_0^a$ | Bias | 0.0442 | -0.0689 | 0.0374 | -0.0193 | -0.0010 | -0.0702 | 0.0299 | -0.0166 |
| | | | | SD | 0.0673 | 0.0401 | 0.0702 | 0.0511 | 0.0561 | 0.0400 | 0.0699 | 0.0511 |
| | | | | RMSE | 0.0806 | 0.0798 | 0.0795 | 0.0547 | 0.0561 | 0.0808 | 0.0760 | 0.0538 |
| (2) | 100 | 10 | $\theta_0^a$ | Bias | 0.0496 | -0.0556 | 0.0244 | -0.0089 | -0.0080 | -0.0580 | 0.0210 | -0.0056 |
| | | | | SD | 0.0467 | 0.0234 | 0.0439 | 0.0326 | 0.0366 | 0.0233 | 0.0440 | 0.0325 |
| | | | | RMSE | 0.0681 | 0.0603 | 0.0502 | 0.0338 | 0.0375 | 0.0625 | 0.0488 | 0.0330 |
| (3) | 100 | 20 | $\theta_0^a$ | Bias | 0.0501 | -0.0252 | 0.0086 | -0.0029 | -0.0088 | -0.0278 | 0.0086 | 0.0003 |
| | | | | SD | 0.0336 | 0.0163 | 0.0306 | 0.0230 | 0.0262 | 0.0163 | 0.0307 | 0.0230 |
| | | | | RMSE | 0.0603 | 0.0300 | 0.0318 | 0.0232 | 0.0276 | 0.0322 | 0.0319 | 0.0230 |
| (4) | 100 | 5 | $\theta_0^b$ | Bias | 0.0468 | -0.1167 | 0.0304 | -0.0397 | 0.0005 | -0.1169 | 0.0365 | -0.0369 |
| | | | | SD | 0.0676 | 0.0448 | 0.0714 | 0.0513 | 0.0562 | 0.0448 | 0.0716 | 0.0513 |
| | | | | RMSE | 0.0822 | 0.1250 | 0.0776 | 0.0648 | 0.0562 | 0.1252 | 0.0803 | 0.0632 |
| (5) | 100 | 10 | $\theta_0^b$ | Bias | 0.0503 | -0.0852 | 0.0103 | -0.0178 | -0.0079 | -0.0863 | 0.0283 | -0.0145 |
| | | | | SD | 0.0495 | 0.0238 | 0.0491 | 0.0328 | 0.0387 | 0.0237 | 0.0483 | 0.0327 |
| | | | | RMSE | 0.0706 | 0.0884 | 0.0502 | 0.0373 | 0.0395 | 0.0895 | 0.0560 | 0.0358 |
| (6) | 100 | 20 | $\theta_0^b$ | Bias | 0.0489 | -0.0393 | -0.0086 | -0.0046 | -0.0100 | -0.0407 | 0.0148 | -0.0013 |
| | | | | SD | 0.0360 | 0.0150 | 0.0379 | 0.0234 | 0.0280 | 0.0150 | 0.0365 | 0.0233 |
| | | | | RMSE | 0.0607 | 0.0421 | 0.0388 | 0.0238 | 0.0297 | 0.0434 | 0.0394 | 0.0234 |
| (7) | 100 | 5 | $\theta_0^c$ | Bias | 0.0431 | -0.1834 | 0.0207 | -0.0857 | -0.0039 | -0.1826 | 0.0396 | -0.0831 |
| | | | | SD | 0.0727 | 0.0478 | 0.0775 | 0.0524 | 0.0599 | 0.0476 | 0.0763 | 0.0523 |
| | | | | RMSE | 0.0845 | 0.1895 | 0.0802 | 0.1004 | 0.0601 | 0.1887 | 0.0859 | 0.0982 |
| (8) | 100 | 10 | $\theta_0^c$ | Bias | 0.0516 | -0.1288 | -0.0063 | -0.0518 | -0.0079 | -0.1288 | 0.0322 | -0.0485 |
| | | | | SD | 0.0511 | 0.0222 | 0.0481 | 0.0332 | 0.0398 | 0.0222 | 0.0440 | 0.0331 |
| | | | | RMSE | 0.0726 | 0.1308 | 0.0485 | 0.0615 | 0.0406 | 0.1307 | 0.0545 | 0.0587 |
| (9) | 100 | 20 | $\theta_0^c$ | Bias | 0.0526 | -0.0608 | -0.0301 | -0.0199 | -0.0072 | -0.0609 | 0.0162 | -0.0166 |
| | | | | SD | 0.0474 | 0.0111 | 0.0415 | 0.0429 | 0.0437 | 0.0111 | 0.0397 | 0.0429 |
| | | | | RMSE | 0.0708 | 0.0618 | 0.0513 | 0.0473 | 0.0442 | 0.0619 | 0.0429 | 0.0460 |

Note: 1. $\theta_0^a$= $(0.2, 0.1, -0.2, 1)$, $\theta_0^b$= $(0.2, 0.5, -0.2, 1)$ and $\theta_0^c$= $(0.2, 0.9, -0.2, 1)$.

2. The IVs are $Y_{n0}, \cdots, Y_{n,t-1}, X_{n1}, \cdots, X_{nT}$ and their first five order spatial lags, for the period $t$.

3. The quadratic matrix is an estimated $I_{T-1} \otimes (\tilde{G}_n - \frac{tr\tilde{G}_n}{n} I_n)$.

Table 7: MLE

| | $n$ | $T$ | $\theta_0$ | | MLE before Bias Correction $\lambda$ | $\gamma$ | $\rho$ | $\beta$ | MLE after Bias Correction $\lambda$ | $\gamma$ | $\rho$ | $\beta$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (1) | 100 | 5 | $\theta_0^a$ | Bias | 0.0043 | -0.1152 | 0.0470 | -0.0267 | 0.0016 | -0.0124 | 0.0061 | -0.0040 |
| | | | | SD | 0.0526 | 0.0357 | 0.0640 | 0.0508 | 0.0531 | 0.0403 | 0.0717 | 0.0509 |
| | | | | RMSE | 0.0528 | 0.1206 | 0.0794 | 0.0573 | 0.0531 | 0.0422 | 0.0720 | 0.0511 |
| (2) | 100 | 10 | $\theta_0^a$ | Bias | 0.0016 | -0.0571 | 0.0218 | -0.0070 | 0.0010 | -0.0041 | 0.0015 | -0.0011 |
| | | | | SD | 0.0359 | 0.0231 | 0.0421 | 0.0324 | 0.0359 | 0.0244 | 0.0442 | 0.0324 |
| | | | | RMSE | 0.0360 | 0.0616 | 0.0475 | 0.0332 | 0.0359 | 0.0248 | 0.0442 | 0.0324 |
| (3) | 100 | 20 | $\theta_0^a$ | Bias | 0.0004 | -0.0275 | 0.0097 | -0.0012 | 0.0002 | -0.0005 | -0.0004 | 0.0003 |
| | | | | SD | 0.0244 | 0.0163 | 0.0297 | 0.0229 | 0.0245 | 0.0167 | 0.0304 | 0.0229 |
| | | | | RMSE | 0.0245 | 0.0319 | 0.0312 | 0.0229 | 0.0245 | 0.0167 | 0.03074 | 0.0229 |
| (4) | 100 | 5 | $\theta_0^b$ | Bias | 0.0060 | -0.1759 | 0.0509 | -0.0556 | 0.0024 | -0.0144 | 0.0022 | -0.0055 |
| | | | | SD | 0.0535 | 0.0376 | 0.0645 | 0.0504 | 0.0548 | 0.0465 | 0.0790 | 0.0519 |
| | | | | RMSE | 0.0538 | 0.1799 | 0.0822 | 0.0751 | 0.0549 | 0.0487 | 0.0791 | 0.0522 |
| (5) | 100 | 10 | $\theta_0^b$ | Bias | 0.0022 | -0.0846 | 0.0233 | -0.0157 | 0.0012 | -0.0062 | 0.0004 | -0.0016 |
| | | | | SD | 0.0363 | 0.0226 | 0.0412 | 0.0326 | 0.0364 | 0.0245 | 0.0446 | 0.0327 |
| | | | | RMSE | 0.0364 | 0.0875 | 0.0474 | 0.0362 | 0.0364 | 0.0253 | 0.0446 | 0.0327 |
| (6) | 100 | 20 | $\theta_0^b$ | Bias | 0.0006 | -0.0401 | 0.0096 | -0.0035 | 0.0003 | -0.0016 | -0.0011 | 0.0002 |
| | | | | SD | 0.0248 | 0.0149 | 0.0291 | 0.0229 | 0.0248 | 0.0154 | 0.0301 | 0.0229 |
| | | | | RMSE | 0.0248 | 0.0428 | 0.0307 | 0.0232 | 0.0248 | 0.0155 | 0.0301 | 0.0229 |
| (7) | 100 | 5 | $\theta_0^c$ | Bias | 0.0009 | -0.2580 | 0.0523 | -0.1179 | -0.0013 | 0.0286 | -0.0114 | 0.0125 |
| | | | | SD | 0.0558 | 0.0383 | 0.0671 | 0.0499 | 0.0626 | 0.0688 | 0.1115 | 0.0600 |
| | | | | RMSE | 0.0558 | 0.2609 | 0.0850 | 0.1280 | 0.0626 | 0.0745 | 0.1121 | 0.0613 |
| (8) | 100 | 10 | $\theta_0^c$ | Bias | 0.0020 | -0.1280 | 0.0252 | -0.0497 | 0.0010 | 0.0092 | -0.0059 | 0.0034 |
| | | | | SD | 0.0371 | 0.0203 | 0.0407 | 0.0328 | 0.0385 | 0.0295 | 0.0522 | 0.0354 |
| | | | | RMSE | 0.0371 | 0.1296 | 0.0478 | 0.0595 | 0.0385 | 0.0309 | 0.0526 | 0.0355 |
| (9) | 100 | 20 | $\theta_0^c$ | Bias | 0.0010 | -0.0610 | 0.0102 | -0.0175 | 0.0000 | 0.0015 | -0.0044 | 0.0010 |
| | | | | SD | 0.0249 | 0.0111 | 0.0273 | 0.0232 | 0.0250 | 0.0137 | 0.0301 | 0.0234 |
| | | | | RMSE | 0.0249 | 0.0620 | 0.0292 | 0.0290 | 0.0250 | 0.0138 | 0.0304 | 0.0235 |

Note: $\theta_0^a$= $(0.2, 0.1, -0.2, 1)$, $\theta_0^b$= $(0.2, 0.5, -0.2, 1)$ and $\theta_0^c$= $(0.2, 0.9, -0.2, 1)$.

those for other estimates are ambiguous. Compared with the 2SLSE in Table 1, the 2SLSE with many IVs (before bias correction) has larger Biases, smaller SDs; for RMSEs, the 2SLSE with many IVs has larger RMSE for most of the estimates when $\gamma_0 = 0.1$ and 0.5 (except for the estimates of $\rho_0$), and has a smaller RMSE for most of the estimates when $\gamma_0 = 0.9$. When we compare the bias corrected 2SLSE using many IVs with the 2SLSE in Table 1, the bias corrected 2SLSE still has larger Biases, but smaller SDs; for the RMSEs, the bias corrected 2SLSE has a smaller RMSE for all the estimates when $\gamma_0 = 0.9$, and a smaller RMSE for the estimates of $\rho_0$. Table 6 is the GMME with many IVs and best quadratic moment. Compared with 2SLSE in Table 5, the GMME has a similar performance, but SDs of the estimates of $\lambda_0$ are smaller. Compared with the BGMME in Table 2, for the items (1)-(5) where we do not have outliers of the estimates in Table 2, the bias corrected GMME has a larger Bias, a smaller SD and the RMSE is larger. By looking at the quantiles of this bias corrected GMME (now listed as the second column block in Table 4) and those of BGMME in Table 3, we see that the bias corrected GMME has a larger Bias, especially a downward Bias for estimate of $\gamma_0$. We also have MLEs before and after bias correction in Table 7. Comparing the MLE with the BGMME in Table 2 and the bias corrected GMME with many IVs and best quadratic moment in Table 6, except that (the bias uncorrected) MLEs of $\gamma_0$ for small $T = 5$ have larger Biases than those of the GMMEs, MLE is slightly better overall, especially when $\gamma_0$ is large.

# 7 Conclusion

This paper proposes the GMM estimation of the spatial dynamic panel data model with fixed effects when $n$ is large and $T$ can be relatively small. We can stack up the data and construct finite moment conditions in a systematic setting, where we derive the best linear and quadratic moment conditions. Alternatively, we can use separate moment conditions, with which the number of IVs may increase as the time period increases. We show that these estimators are $\sqrt{nT}$ consistent, asymptotically normal, and have efficient properties.

In a simple dynamic panel data model with fixed effects, the OLS (least squares with dummy variables; within) estimate has $O(1/T)$ bias due to the correlation of predetermined variables and resulting disturbances after the elimination of fixed effects. The IV estimation approach avoids such a problem when a finite number of IVs is used as those IVs are uncorrelated with the disturbances. However, when the number of IVs increases without bound as the sample size increases, the correlation of the predetermined variables and disturbances is restored to some extent (determined by the number of IVs). In the SDPD model with fixed effects, and time and spatial time lags, the OLS estimate has a similar $O(1/T)$ bias (Korniotis 2008). For the SDPD model with the additional contemporaneous spatial lag, an additional $O(1)$ bias for the OLS estimate occurs. The latter is due to the simultaneity of the spatial lag variable. The simultaneity of the spatial lag can be handled in the QML approach as in Yu et al. (2008); but, the bias order $O(1/T)$ remains for the QMLE. On the contrary, IV estimates would not have such an order of bias when the number of IVs is finite. However, when the number of IVs increases (to infinity), the bias for the SDPD model will also be restored, and the bias for the estimate of the spatial effect, $\varphi_1$, would be dominant. A bias correction procedure can eliminate this dominating bias. Therefore, under the situation that $T$ is small relative to $n$, we can have consistent estimates with properly centered asymptotic normal distribution.

In addition to linear moments constructed from the time lags, spatial time lags, and exogenous variables, we also utilize quadratic moments to increase the efficiency of the estimates. These quadratic moments are implied by the spatial effect in the SDPD model, which do not appear in the dynamic panel data models. This is a distinct feature of our GMM approach as compared with IV approaches for the estimation of spatial dynamic models. We propose an optimal quadratic moment condition that is free of distributional assumption for the disturbances.

The best GMM estimates from the finite moment conditions in the systematic setting have the same asymptotic distribution of the MLE when the disturbances are normal. Compared to MLE of the SDPD models, the GMM estimate is computationally simpler, and can be extended as in Lee and Liu (2010) to higher spatial order models that the MLE cannot easily deal with. Additionally, when the distribution is not normal, the best GMM estimate in the current paper can be more efficient relative to the QMLE as the kurtosis of the disturbances is used for the best quadratic moment. The many moment approach and the MLE approach complement each other as the former can be applied to the case with $T$ being small relative to $n$, while the ML approach is valid for the case with $T$ being moderate or large relative to $n$.

# Appendices

## A Notations

The following list summarizes some frequently used notations in the paper:

$S_n(\lambda) = I_n - \lambda W_n$ for any possible $\lambda$, $S_n = I_n - \lambda_0 W_n$, $G_n = W_n S_n^{-1}$ and $A_n = S_n^{-1}(\gamma_0 I_n + \rho_0 W_n)$.

$\mathcal{A}_n^s = \mathcal{A}_n' + \mathcal{A}_n$ for any square matrix $\mathcal{A}_n$.

$vec_D(\mathcal{A}_n)$ is the column vector formed by diagonal elements of $\mathcal{A}_n$.

$F_{T,T-1}$ is the $T \times (T-1)$ matrix of Helmert transformation.

$F_{n,n-1}$ is the $n \times (n-1)$ eigenvectors matrix of $J_n = I_n - \frac{1}{n} l_n l_n'$ corresponding to the eigenvalues of one.

$[Y_{n1}^*, \cdots, Y_{n,T-1}^*] = [Y_{n1}, \cdots, Y_{nT}]F_{T,T-1}$, $[Y_{n0}^{(*,-1)}, \cdots, Y_{n,T-2}^{(*,-1)}] = [Y_{n0}, \cdots, Y_{n,T-1}]F_{T,T-1}$.

$V_{nt}^* = (\frac{T-t}{T-t+1})^{\frac{1}{2}}[V_{nt} - \frac{1}{T-t}\sum_{h=t+1}^{T} V_{nh}]$.

$Z_{nt} = (Y_{n,t-1}, W_n Y_{n,t-1}, X_{nt})$ and $Z^*_{nt} = (Y^{(*,-1)}_{n,t-1}, W_n Y^{(*,-1)}_{n,t-1}, X^*_{nt})$.

$\theta = (\lambda, \delta')'$ with $\delta = (\gamma, \rho, \beta')$.

$[Y^{**}_{n-1,1}, \cdots, Y^{**}_{n-1,,T-1}] = F'_{n,n-1}[Y^*_{n1}, \cdots, Y^*_{n,T-1}]$, $[Y^{(**,-1)}_{n-1,0}, \cdots, Y^{(**,-1)}_{n-1,T-2}] = F'_{n,n-1}[Y^{(*,-1)}_{n0}, \cdots, Y^{(*,-1)}_{n,T-2}]$.

$k_x$ is the column dimension of $X_{nt}$ and $k_z = k_x + 2$ is the column dimension of $Z^*_{nt}$.

$\mathcal{P}_n$ is the class of $n \times n$ nonstochastic matrix with a zero trace.

$Q_{nt}$ is the IV matrix for Section 3, $H_{nt} = (h_{nt}, W_n h_{nt}, \cdots, W_n^{p_n} h_{nt})$ is the IV matrix for Section 4 where $h_{nt} = [Y_{n0}, ..., Y_{n,t-1}, X_{n1}, ..., X_{nT}]$, and $M_{nt} = H_{nt}(H'_{nt}H_{nt})^+ H'_{nt}$.

$\mathbf{Z}^*_{n,T-1} = (Z^{*\prime}_{n1}, ..., Z^{*\prime}_{n,T-1})'$ and $\mathbf{V}^*_{n,T-1} = (V'^*_{n1}, ..., V'^*_{n,T-1})'$.

$\mathbf{W}_{n,T-1} = I_{T-1} \otimes W_n$, $\mathbf{S}_{n,T-1} = I_{T-1} \otimes S_n$, $\mathbf{G}_{n,T-1} = I_{T-1} \otimes G_n$, and $\mathbf{Q}_{n,T-1} = (Q'_{n1}, \cdots, Q'_{n,T-1})'$.

$\mathbf{P}_{n,T-1,l} = I_{T-1} \otimes P_{nl}$ for $l = 1, 2, ...m$, where $P_{nl}$ is from $\mathcal{P}_n$ in Sections 3 and 4.

$\mathbf{J}_{n,T-1} = I_{T-1} \otimes J_n$.

$\omega_{nm,T} = [vec_D(\mathbf{P}_{n,T-1,1}), \cdots, vec_D(\mathbf{P}_{n,T-1,m})]$ and $\omega^s_{nm,T} = [vec_D(\mathbf{P}^s_{n,T-1,1}), \cdots, vec_D(\mathbf{P}^s_{n,T-1,m})]$.

$\Phi_j = \sum_{h=0}^{j-1} A_n^h$, $\Psi_t = c_{Tt}\left(I_n - \frac{A_n \Phi_{T-t}}{T-t}\right)$ where $c_{Tt} = (\frac{T-t}{T-t+1})^{\frac{1}{2}}$.

$\tilde{X}_{n,tT} = \frac{1}{T-t} S_n^{-1} \sum_{h=t}^{T-1} \Phi_{T-h} X_{nh}$ and $\tilde{V}_{n,tT} = \frac{1}{T-t} S_n^{-1} \sum_{h=t}^{T-1} \Phi_{T-h} V_{nh}$.

$E(Y^{(*,-1)}_{n,t-1}|\mathcal{I}_{t-1}) = \Psi_t Y^w_{n,t-1} - c_{Tt}\tilde{X}_{n,tT}\beta_0$ where $Y^w_{n,t-1} = Y_{n,t-1} - (I_n - A_n)^{-1} S_n^{-1}\mathbf{c}_{n0}$.

$\mathbb{H}_{n1} = \Psi_1 Y_{n0} - c_{T1}\tilde{X}_{1T}\beta_0$ and $\mathbb{H}_{nt}$ is in (11) for $t \geq 2$.

$\mathbb{K}_{nt} \equiv (\mathbb{H}_{nt}, W_n\mathbb{H}_{nt}, X^*_{nt})$ and $\mathbb{Q}_{nt} = (G_n\mathbb{K}_{nt}\delta_0, \mathbb{K}_{nt})$.

$\mathbf{\Sigma}_{nT,22} = \frac{1}{n(T-1)}(\mathbf{G}_{n,T-1}\mathbf{Z}^*_{n,T-1}\delta_0, \mathbf{Z}^*_{n,T-1})'(\mathbf{G}_{n,T-1}\mathbf{Z}^*_{n,T-1}\delta_0, \mathbf{Z}^*_{n,T-1})$.

$f_{nt} = [G_n((\gamma_0 I_n + \rho_0 W_n)E(Y^{(*,-1)}_{n,t-1}|\mathcal{I}_{t-1}) + X^*_{nt}\beta_0), E(Y^{(*,-1)}_{n,t-1}|\mathcal{I}_{t-1}), W_n E(Y^{(*,-1)}_{n,t-1}|\mathcal{I}_{t-1}), X^*_{nt}]$.

$u_{nt} = [G_n((\gamma_0 I_n + \rho_0 W_n)\eta_{nt} + V^*_{nt}), \eta_{nt}, W_n\eta_{nt}, \mathbf{0}_{n\times k_x}]$ with $\eta_{nt} = -c_{Tt}\tilde{V}_{n,tT}$.

# B Some Lemmas

**Lemma 1** *Under Assumption 2, for any $n \times n$ nonstochastic UB matrices $B_n$,*

*(i) $E(V^*_{nt}B_nV^{*\prime}_{ns}|\mathcal{I}_{t-1}) = 0$ for $t \neq s$;*

*(ii) $\frac{1}{n(T-1)}\mathbf{V}^{*\prime}_{n,T-1}(I_{T-1}\otimes B_n)\mathbf{V}^*_{n,T-1} = \frac{1}{n}\sigma_0^2 trB_n + O_p\left(\frac{1}{\sqrt{nT}}\right)$;*

*(iii) under Assumption 6, $\frac{1}{n(T-1)}\mathbf{Y}^{(*,-1)\prime}_{n,T-1}(I_{T-1}\otimes B_n)\mathbf{V}^*_{n,T-1} - E\frac{1}{n(T-1)}\mathbf{Y}^{(*,-1)\prime}_{n,T-1}(I_{T-1}\otimes B_n)\mathbf{V}^*_{n,T-1} = O_p\left(\frac{1}{\sqrt{nT}}\right)$ where $E\frac{1}{n(T-1)}\mathbf{Y}^{(*,-1)\prime}_{n,T-1}(I_{T-1}\otimes B_n)\mathbf{V}^*_{n,T-1} = O\left(\frac{1}{T}\right)$;*

*(iv) under Assumption 9, for the IV matrix $Q_{nt}$, $\text{plim}_{n\to\infty}\frac{1}{n(T-1)}\sum_{t=1}^{T-1}Q'_{nt}B_nV^*_{nt} = 0$.*

*(v) $E[(V^{*\prime}_{nt}B_nV^*_{nt})^2] = (\mu_4 - 3\sigma_0^4)c^4_{Tt}\left(1 + \frac{1}{(T-t)^3}\right)vec'_D(B_n)vec_D(B_n) + \sigma_0^4[tr^2(B_n) + tr(B_nB_n^s)]$.*

**Lemma 2** *Under Assumption 2 with $\mathbf{P}_{n,T-1,j} = I_{T-1}\otimes P_{nj}$, the covariance of $\mathbf{V}^{*\prime}_{n,T-1}\mathbf{P}_{n,T-1,j}\mathbf{V}^*_{n,T-1}$ and $\mathbf{V}^{*\prime}_{n,T-1}\mathbf{P}_{n,T-1,l}\mathbf{V}^*_{n,T-1}$ is $\sigma_0^4 tr(\mathbf{P}_{n,T-1,j}\mathbf{P}^s_{n,T-1,l}) + (\mu_4 - 3\sigma_0^4)vec'_D(\mathbf{P}_{n,T-1,j})vec_D(\mathbf{P}_{n,T-1,l})$ and that of $\mathbf{V}^{*\prime}_{n,T-1}\mathbf{P}_{n,T-1,j}\mathbf{V}^*_{n,T-1}$ and $\mathbf{Q}'_{n,T-1}\mathbf{V}^*_{n,T-1}$ is zero for $j = 1, ..., m$.*

Let $C_{nt}$ be an $n \times 1$ column vector from the IV matrix $Q_{nt}$ in Assumption 9. Denote $\boldsymbol{s}_{nt} = C'_{nt}V^*_{nt} + V^{*\prime}_{nt}B_nV^*_{nt} - \sigma_0^2 trB_n$ and[16] $\sigma^2_{\mathbf{s},nT} = \mathrm{Var}(\sum_{t=1}^{T-1}\boldsymbol{s}_{nt}) = E(\sigma_0^2\sum_{t=1}^{T-1}C'_{nt}C_{nt} + T(\mu_4 - 3\sigma_0^4)\sum_{i=1}^n b^2_{n,ii} + T\sigma_0^4 tr(B_nB_n^s))$.

**Lemma 3** *Under Assumptions 2, 8 and 9, if* $\{\frac{1}{n(T-1)}\sigma^2_{\mathbf{s},nT}\}$ *is bounded away from zero,* $\frac{\sum_{t=1}^{T-1}\boldsymbol{s}_{nt}}{\sigma_{\mathbf{s},nT}} \xrightarrow{d} N(0,1)$.

Given square matrices $P_{nl}$, $l = 1, ..., m$, with zero trace, where the quadratic moments in (3) and (4) take the form $\mathbf{P}_{n,T-1,l} = I_{T-1} \otimes P_{nl}$, let $\eta_4 = \frac{\mu_4}{\sigma_0^4}$, $P^+_{nl} = P_{nl} - diag(P_{nl}) + \sqrt{\frac{\eta_4 - 1}{2}}diag(P_{nl})$ and

$$G_n^- = G_n - \frac{trG_n}{n}I_n + \left(\sqrt{\frac{2}{\eta_4 - 1}} - 1\right)\left(diag(G_n) - \frac{trG_n}{n}I_n\right),$$

so that $\mathbf{P}^+_{n,T-1,l} = I_{T-1} \otimes P^+_{nl}$ and $\mathbf{G}^-_{n,T-1} = I_{T-1} \otimes G^-_n$. Denote $\Sigma_{P,nT} = \frac{1}{n(T-1)}C_{mn,T}(\frac{\mu_4 - 3\sigma_0^4}{\sigma_0^4}\omega'_{nm,T}\omega_{nm,T} + \Delta_{mn,T})^{-1}C'_{mn,T}$ in (7) where $C_{mn,T} = [tr(\mathbf{P}^s_{n,T-1,1}\mathbf{G}_{n,T-1}), \ldots, tr(\mathbf{P}^s_{n,T-1,m}\mathbf{G}_{n,T-1})]$.

**Lemma 4** *(i)* $tr(\mathbf{P}^s_{n,T-1,l}\mathbf{G}_{n,T-1}) = tr(\mathbf{P}^{+s}_{n,T-1,l}\mathbf{G}^-_{n,T-1})$;

*(ii)* $\frac{\mu_4 - 3\sigma_0^4}{\sigma_0^4}\omega'_{nm,T}\omega_{nm,T} + \Delta_{mn,T} = \frac{1}{2}(vec(\mathbf{P}^{+s}_{n,T-1,1}), ..., vec(\mathbf{P}^{+s}_{n,T-1,m}))'(vec(\mathbf{P}^{+s}_{n,T-1,1}), ..., vec(\mathbf{P}^{+s}_{n,T-1,m}))$;

*(iii)* $\Sigma_{P,nT} \leq \frac{1}{2}vec'(\mathbf{G}^{-s}_{n,T-1})vec(\mathbf{G}^{-s}_{n,T-1})$;

*(iv) for* $\mathbf{P}_{n,T-1}$ *in (8), we have* $tr(\mathbf{P}^{+s}_{n,T-1}\mathbf{G}^-_{n,T-1})[\frac{1}{2}(vec'(\mathbf{P}^{+s}_{n,T-1})'vec(\mathbf{P}^{+s}_{n,T-1})]^{-1}tr(\mathbf{P}^{+s}_{n,T-1}\mathbf{G}^-_{n,T-1}) = \frac{1}{2}vec'(\mathbf{G}^{-s}_{n,T-1})vec(\mathbf{G}^{-s}_{n,T-1})$*, where* $\frac{1}{2}vec'(\mathbf{G}^{-s}_{n,T-1})vec(\mathbf{G}^{-s}_{n,T-1}) = tr(\mathbf{P}^s_{n,T-1}\mathbf{G}_{n,T-1})$.

**Lemma 5** *Denote* $c_{Tt} = (\frac{T-t}{T-t+1})^{\frac{1}{2}}$ *and* $\Phi_j = \sum_{h=0}^{j-1}A^h_n$*. We have* $Y^{(*,-1)}_{n,t-1} = E(Y^{(*,-1)}_{n,t-1}|\mathcal{I}_{t-1}) + \eta_{nt}$ *where*

*(i)* $E(Y^{(*,-1)}_{n,t-1}|\mathcal{I}_{t-1}) = c_{Tt}(I_n - \frac{A_n\Phi_{T-t}}{T-t})(Y_{n,t-1} - (I_n - A_n)^{-1}S_n^{-1}\mathbf{c}_{n0}) - c_{Tt}\tilde{X}_{n,tT}\beta_0$*, and*

*(ii)* $\eta_{nt} = -c_{Tt}\tilde{V}_{n,tT}$ *with* $\tilde{V}_{n,tT} = \frac{1}{T-t}S_n^{-1}\sum_{h=t}^{T-1}\Phi_{T-h}V_{nh}$.

*For* $\mathbb{H}_{nt}$ *in (11),*

$$E(Y^{(*,-1)}_{n,t-1}|\mathcal{I}_{t-1}) = \mathbb{H}_{nt} + \mathbb{W}_{nt}, \text{ and } Y^{(*,-1)}_{n,t-1} = \mathbb{H}_{nt} + \mathbb{W}_{nt} + \eta_{nt}, \quad (32)$$

*where* $\mathbb{W}_{n1} = -\Psi_1(I_n - A_n)^{-1}S_n^{-1}\mathbf{c}_{n0}$ *and* $\mathbb{W}_{nt} = \Psi_t(I_n - A_n)^{-1}S_n^{-1}\frac{1}{t-1}\sum_{s=1}^{t-1}V_{ns}$ *for* $t \geq 2$.

**Lemma 6** *Under Assumptions 1-8 and* $T \to \infty$*, for any nonstochastic square matrix* $B_n$*,*

$$\frac{1}{n(T-1)}\sum_{t=1}^{T-1}\mathbb{H}'_{nt}B_n\mathbb{H}_{nt} = E(\frac{1}{n(T-1)}\sum_{t=1}^{T-1}\mathbb{H}'_{nt}B_n\mathbb{H}_{nt}) + O_p\left(\frac{1}{\sqrt{nT}}\right)$$

*with* $E(\frac{1}{n(T-1)}\sum_{t=1}^{T-1}\mathbb{H}'_{nt}B_n\mathbb{H}_{nt}) = O(1)$*. Also,*

$$\frac{1}{n(T-1)}\sum_{t=1}^{T-1}\mathbb{W}'_{nt}B_n\mathbb{W}_{nt} = o_p(1) \text{ and } \frac{1}{n(T-1)}\sum_{t=1}^{T-1}\mathbb{H}'_{nt}B_n\mathbb{W}_{nt} = o_p(1). \quad (33)$$

*Similarly,*

$$\frac{1}{n(T-1)}\sum_{t=1}^{T-1}(\mathbb{W}_{nt} + \eta_{nt})'B_n(\mathbb{W}_{nt} + \eta_{nt}) = o_p(1) \text{ and } \frac{1}{n(T-1)}\sum_{t=1}^{T-1}\mathbb{H}'_{nt}B_n(\mathbb{W}_{nt} + \eta_{nt}) = o_p(1). \quad (34)$$

[16] Here, the covariance of $\sum_{t=1}^{T-1}C'_{nt}V^*_{nt}$ and $\sum_{t=1}^{T-1}V^*_{nt}B_nV^*_{nt}$ is zero, which is similar to Lemma 2. Thus, the third moment of $v_{it}$ does not appear in $\sigma^2_{\mathbf{s},nT}$.

Let $M_{nt} = H_{nt}(H'_{nt}H_{nt})^{+}H'_{nt}$ so that $M_{nt}$ is an $n \times n$ idempotent matrix with rank $K_t$.

**Lemma 7** *For any UB $n \times n$ square matrices $B_{1n}$ and $B_{2n}$,*

*(i) $tr(M_{nt}B_{1n}B'_{1n}M_{nt}) \leq cK_t$, where $c$ is a finite constant (for all $n$ and $t$);*

*(ii) $|tr(B_{1n}M_{nt})|$ and $|tr(B_{1n}M_{nt}B_{2n})|$ are less than $cK_t$ for some $c > 0$;*

*(iii) $|tr(M_{nt}B_{1n}M_{ns}B_{2n})|$ are less than $c\sqrt{K_tK_s}$ for some $c > 0$.*

**Lemma 8** *Under Assumptions 1-6, for any nonstochastic UB matrix $B_n$,*

*(i) $E(\sum_{t=1}^{T-1} V^{*\prime}_{nt}B_nM_{nt}V^*_{nt}) = \sigma_0^2\sum_{t=1}^{T-1}E[tr(B_nM_{nt})] = O(\sum_{t=1}^{T-1}K_t)$, and*

*(ii) $\sum_{t=1}^{T-1}(V^{*\prime}_{nt}B_nM_{nt}V^*_{nt} - \sigma_0^2 tr(B_nM_{nt})) = O_p(\sum_{t=1}^{T-1}\sqrt{K_t})$.*

**Lemma 9** *Under Assumptions 1-6, for any nonstochastic UB matrix $B_n$,*

$$E(\sum_{t=1}^{T-1}\eta'_{nt}B_nM_{nt}V^*_{nt}) = -\frac{\sigma_0^2}{T+1-t}E[tr(M_{nt}C'_{nTt}S_n^{'-1}B'_n)] = O\left(\sum_{t=1}^{T-1}\frac{K_t}{(T+1-t)(T-t)}\right),$$

*and $\sum_{t=1}^{T-1}(\eta'_{nt}B_nM_{nt}V^*_{nt} + \frac{\sigma_0^2}{T+1-t}tr(M_{nt}C'_{nTt}S_n^{'-1}B'_n)) = O_p\left(\sum_{t=1}^{T-1}\sqrt{\frac{K_t}{T+1-t}}\right)$, where $C_{nTt} = \frac{1}{T-t}(I_n + 2A_n + \cdots + (T-t)A_n^{T-1-t})$. Also, for any nonstochastic UB matrices $B_{n1}$ and $B_{n2}$, $E(\sum_{t=1}^{T-1}\eta'_{nt}B_{n1}M_{nt}B_{n2}\eta_{nt}) = O\left(\sum_{t=1}^{T-1}\frac{K_t}{T-t+1}\right)$ and $\sum_{t=1}^{T-1}(\eta'_{nt}B_{n1}M_{ntBn2}\eta_{nt} - E(\eta'_{nt}B_{n1}M_{nt}B_{n2}\eta_{nt}|\mathcal{I}_{t-1})) = O_p\left(\sum_{t=1}^{T-1}\frac{\sqrt{K_t}}{T-t+1}\right)$.*

**Lemma 10** *Under Assumptions 1-8 , suppose we choose $H_{nt}$ from (16). For each $t$, there exists a matrix $\boldsymbol{\pi}_t$ such that $\frac{1}{n(T-1)}\sum_{t=1}^{T-1}(f_{nt} - H_{nt}\cdot\boldsymbol{\pi}_t)'(f_{nt} - H_{nt}\cdot\boldsymbol{\pi}_t) \xrightarrow{p} 0$ as $n \to \infty$ and $T \to \infty$.*

The following Lemma 11 is about magnitudes of certain orders in the 2SLS estimate with many IVs in (19). Denote $K = \max\{K_1, ..., K_{T-1}\}$, $e_f(K) = \frac{1}{n(T-1)}\sum_{t=1}^{T-1}f'_{nt}(I_n - M_{nt})f_{nt}$ and $\Delta_K = tr(e_f(K))$.

**Lemma 11** *Under Assumptions 1-8 and $T \to \infty$,*

*(i) $\Delta_K = o_p(1)$;*

*(ii) $\frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^{T-1}f'_{nt}(I_n - M_{nt})V^*_{nt} = O_p((E\Delta_K)^{1/2})$;*

*(iii) $\frac{1}{n(T-1)}\sum_{t=1}^{T-1}f'_{nt}M_{nt}B_nV^*_{nt} = O_p\left(\frac{1}{\sqrt{n(T-1)}}\right)$ and $\frac{1}{n(T-1)}\sum_{t=1}^{T-1}f'_{nt}M_{nt}B_n\tilde{V}_{n,tT} = O_p\left(\frac{1}{\sqrt{n(T-1)}}\right)$;*

*(iv) $\frac{1}{n(T-1)}\sum_{t=1}^{T-1}(u'_{nt}M_{nt}u_{nt} - E(u'_{nt}M_{nt}u_{nt}|\mathcal{I}_{t-1})) = O_p(\frac{1}{n(T-1)}\sum_{t=1}^{T-1}\sqrt{K_t})$ where $\frac{1}{n(T-1)}\sum_{t=1}^{T-1}E(u'_{nt}M_{nt}u_{nt}|\mathcal{I}_{t-1}) = O(\frac{1}{n(T-1)}\sum_{t=1}^{T-1}K_t)$.*

**Proof for Lemma 1:**

(i) As $V^*_{nt}B_nV^{*\prime}_{ns} = c_{Tt}c_{Ts}(V_{nt} - \frac{1}{T-t}\sum_{h=t+1}^{T}V_{nh})B_n(V_{ns} - \frac{1}{T-s}\sum_{h=s+1}^{T}V_{nh})'$, for $t > s$,

$$\begin{aligned}
E(V^*_{nt}B_nV^{*\prime}_{ns}|\mathcal{I}_{t-1}) &= c_{Tt}c_{Ts}E[(V_{nt} - \frac{1}{T-t}\sum_{h=t+1}^{T}V_{nh})B_n(V_{ns} - \frac{1}{T-s}\sum_{h=s+1}^{t-1}V_{nh})'|\mathcal{I}_{t-1}] \\
&\quad - c_{Tt}c_{Ts}E[(V_{nt} - \frac{1}{T-t}\sum_{h=t+1}^{T}V_{nh})B_n\frac{1}{T-s}\sum_{h=t}^{T}V'_{nh}|\mathcal{I}_{t-1}] \\
&= -c_{Tt}c_{Ts}\sigma_0^2[\frac{1}{T-s}tr(B_n) - \frac{1}{(T-t)}\frac{1}{(T-s)}(T-t)tr(B_n)] = 0.
\end{aligned}$$

For $t < s$, we have $E(V_{nt}^* B_n V_{ns}^{*\prime} | \mathcal{I}_{t-1}) = E(V_{nt}^* B_n V_{ns}^{*\prime}) = 0$.

(ii) We have $\frac{1}{n(T-1)} \mathbf{V}_{n,T-1}^{*\prime} (I_{T-1} \otimes B_n) \mathbf{V}_{n,T-1}^* = \frac{1}{n(T-1)} \sum_{t=1}^T \tilde{V}_{nt}' B_n \tilde{V}_{nt}$ where $\tilde{V}_{nt} = V_{nt} - \frac{1}{T} \sum_{s=1}^T V_{ns}$. By Lemma 9 in Yu et al. (2008), $\frac{1}{n(T-1)} \sum_{t=1}^T \tilde{V}_{nt}' B_n \tilde{V}_{nt} - E \frac{1}{n(T-1)} \sum_{t=1}^T \tilde{V}_{nt}' B_n \tilde{V}_{nt} = O_p\left(\frac{1}{\sqrt{nT}}\right)$ with $E \frac{1}{n(T-1)} \sum_{t=1}^T \tilde{V}_{nt}' B_n \tilde{V}_{nt} = \frac{1}{n} \sigma_0^2 tr B_n$.

(iii) We have $\frac{1}{n(T-1)} \mathbf{Y}_{n,T-1}^{(*,-1)\prime} (I_{T-1} \otimes B_n) \mathbf{V}_{n,T-1}^* = \frac{1}{n(T-1)} \sum_{t=1}^T \tilde{Y}_{n,t-1}' B_n \tilde{V}_{nt}$, where $\tilde{Y}_{n,t-1} = Y_{n,t-1} - \frac{1}{T} \sum_{s=0}^{T-1} Y_{ns}$. Under Assumption 6, $\frac{1}{n(T-1)} \sum_{t=1}^T \tilde{Y}_{n,t-1}' B_n \tilde{V}_{nt} - E \frac{1}{n(T-1)} \sum_{t=1}^T \tilde{Y}_{n,t-1}' B_n \tilde{V}_{nt} = O_p\left(\frac{1}{\sqrt{nT}}\right)$ by Lemma 15 in Yu et al. (2008), with $E \frac{1}{n(T-1)} \sum_{t=1}^T \tilde{Y}_{n,t-1}' B_n \tilde{V}_{nt} = O\left(\frac{1}{T}\right)$.

(iv) From Assumption 9 that $E(Q_{nt} | I_{t-1}) = Q_{nt}$, $E(\frac{1}{n(T-1)} \sum_{t=1}^{T-1} Q_{nt}' B_n V_{nt}^*) = 0$. Also, as $E(V_{nt}^* V_{ns}^{*\prime} | \mathcal{I}_{t-1}) = 0$ whenever $s < t$, $\mathrm{Cov}(Q_{nt}' B_n V_{nt}^*, Q_{ns}' B_n V_{ns}^*) = 0$ for $t \neq s$. Hence, $\mathrm{Var}(\frac{1}{n(T-1)} \sum_{t=1}^{T-1} Q_{nt}' B_n V_{nt}^*) = \frac{\sigma_0^2}{n^2(T-1)^2} \sum_{t=1}^{T-1} E[Q_{nt}' B_n B_n' Q_{nt}]$. Under Assumption 9 that elements in $Q_{nt}$ are $O_p(1)$ uniformly in $n$ and $t$, $\frac{1}{n(T-1)} \sum_{t=1}^{T-1} E[Q_{nt}' B_n B_n' Q_{nt}] = O(1)$; hence, $\frac{1}{n(T-1)} \sum_{t=1}^{T-1} Q_{nt}' B_n V_{nt}^* = O_p\left(\frac{1}{\sqrt{n(T-1)}}\right)$.

(v) As $V_{nt}^* = c_{Tt}[(1, -\frac{1}{T-t}, ..., -\frac{1}{T-t}) \otimes I_n](V_{nt}', ..., V_{nT}')'$, we have $V_{nt}^{*\prime} B_n V_{nt}^* = c_{Tt}^2 (V_{nt}', ..., V_{nT}') A_{nT} (V_{nt}', ..., V_{nT}')'$ where $A_{nT} = [(1, -\frac{1}{T-t}, ..., \frac{1}{T-t})' \otimes I_n] B_n [(1, -\frac{1}{T-t}, ..., -\frac{1}{T-t}) \otimes I_n]$. It follows that $tr(A_{nT}) = \frac{T-t+1}{T-t} tr(B_n)$, $tr(A_{nT}' A_{nT}^s) = (\frac{T-t+1}{T-t})^2 tr(B_n B_n^s)$ and $vec_D'(A_{nT}) vec_D(A_{nT}) = (1 + \frac{1}{(T-t)^3}) vec_D'(B_n) vec_D(B_n)$. Hence, as $c_{Tt}^2 = \frac{T-t}{T-t+1}$, we have the result in (v). ■

**Proof for Lemma 2:** Denote $\mathbf{V}_{nT} = (V_{n1}', \cdots, V_{nT}')'$. As $\mathbf{V}_{n,T-1}^* = (F_{T,T-1}' \otimes I_n) \mathbf{V}_{nT}$ and $\mathbf{P}_{n,T-1,j} = I_{T-1} \otimes P_{nj}$, we have $\mathrm{Cov}(\mathbf{V}_{n,T-1}^{*\prime} \mathbf{P}_{n,T-1,j} \mathbf{V}_{n,T-1}^*, \mathbf{V}_{n,T-1}^{*\prime} \mathbf{P}_{n,T-1,l} \mathbf{V}_{n,T-1}^*) = \mathrm{Cov}(\mathbf{V}_{nT}' (J_T \otimes P_{nj}) \mathbf{V}_{nT}, \mathbf{V}_{nT}' (J_T \otimes P_{nl}) \mathbf{V}_{nT}) = \sigma_0^4 tr((J_T \otimes P_{nj})(J_T \otimes P_{nl}^s)) + (\mu_4 - 3\sigma_0^4) vec_D'(J_T \otimes P_{nj}) vec_D(J_T \otimes P_{nl})$, by using the variance formulae of quadratic form of *i.i.d.* disturbances. Using $tr((J_T \otimes P_{nj})(J_T \otimes P_{nl}^s)) = (T-1) tr(P_{nj} P_{nl}^s) = tr(\mathbf{P}_{n,T-1,j} \mathbf{P}_{n,T-1,l}^s)$ and $vec_D'(J_T \otimes P_{nj}) vec_D(J_T \otimes P_{nl}) = (T-1) vec_D'(P_{nj}) vec_D(P_{nl}) = vec_D'(\mathbf{P}_{n,T-1,j}) vec_D(\mathbf{P}_{n,T-1,l})$, the covariance matrix of $\mathbf{V}_{n,T-1}^{*\prime} \mathbf{P}_{n,T-1,j} \mathbf{V}_{n,T-1}^*$ and $\mathbf{Q}_{n,T-1}' \mathbf{V}_{n,T-1}^*$ is $\sigma_0^4 tr(\mathbf{P}_{n,T-1,j} \mathbf{P}_{n,T-1,l}^s) + (\mu_4 - 3\sigma_0^4) vec_D'(\mathbf{P}_{n,T-1,j}) vec_D(\mathbf{P}_{n,T-1,l})$.

For $\mathrm{Cov}(\mathbf{V}_{n,T-1}^{*\prime} \mathbf{P}_{n,T-1,j} \mathbf{V}_{n,T-1}^*, \mathbf{Q}_{n,T-1}' \mathbf{V}_{n,T-1}^*) = E[(\sum_{t=1}^{T-1} V_{nt}^{*\prime} P_{n,j} V_{nt}^*)(\sum_{t=1}^{T-1} Q_{nt}' V_{nt}^*)]$, we have

$$
\begin{aligned}
& E(V_{nt}^{*\prime} P_{n,j} V_{nt}^*)(Q_{nt}' V_{nt}^*) \\
= \; & c_{Tt}^3 E\left[(V_{nt} - \frac{1}{T-t} \sum_{h=t+1}^T V_{nh})' P_{n,j} (V_{nt} - \frac{1}{T-t} \sum_{h=t+1}^T V_{nh}) Q_{nt}' (V_{nt} - \frac{1}{T-t} \sum_{h=t+1}^T V_{nh})\right] \\
= \; & c_{Tt}^3 \mu_3 E Q_{nt}' vec_D(P_{n,j})(1 - \frac{1}{(T-t)^2}) = \mu_3 E Q_{nt}' vec_D(P_{n,j}) \cdot c_{Tt}(1 - \frac{1}{T-t}).
\end{aligned}
$$

For $s < t$, we have

$$
\begin{aligned}
& E(V_{ns}^{*\prime}P_{n,j}V_{ns}^*)(Q_{nt}'V_{nt}^*) \\
= \; & c_{Ts}^2 c_{Tt} E\left[(V_{ns} - \frac{1}{T-s}\textstyle\sum_{h=s+1}^T V_{nh})'P_{n,j}(V_{ns} - \frac{1}{T-s}\sum_{h=s+1}^T V_{nh})Q_{nt}'(V_{nt} - \frac{1}{T-t}\sum_{h=t+1}^T V_{nh})\right] \\
= \; & c_{Ts}^2 c_{Tt} E\left[(-\frac{1}{T-s}\textstyle\sum_{h=t}^T V_{nh})'P_{n,j}(-\frac{1}{T-s}\sum_{h=t}^T V_{nh})Q_{nt}'(V_{nt} - \frac{1}{T-t}\sum_{h=t+1}^T V_{nh})\right] \\
= \; & c_{Ts}^2 c_{Tt}\mu_3 E Q_{nt}' vec_D(P_{n,j})[\frac{1}{(T-s)^2} - \frac{T-t}{(T-s)^2(T-t)}] = 0,
\end{aligned}
$$

because $EV_{ng}'P_{n,j}V_{nh}Q_{nt}'V_{np} = 0$ for $g, h < t$ and $p \geq t$, $E[V_{nt} - \frac{1}{T-t}\sum_{h=t+1}^T V_{nh}](\sum_{h=t}^T V_{nh}) = 0$ and $E(Q_{nt}|\mathcal{I}_{t-1}) = Q_{nt}$. For $s < t$, we have

$$
\begin{aligned}
& E(V_{nt}^{*\prime}P_{n,j}V_{nt}^*)(Q_{ns}'V_{ns}^*) \\
= \; & c_{Tt}^2 c_{Ts} E\left[(V_{nt} - \frac{1}{T-t}\textstyle\sum_{h=t+1}^T V_{nh})'P_{n,j}(V_{nt} - \frac{1}{T-t}\sum_{h=t+1}^T V_{nh})Q_{ns}'(V_{ns} - \frac{1}{T-s}\sum_{h=s+1}^T V_{nh})\right] \\
= \; & c_{Tt}^2 c_{Ts} E\left[(V_{nt} - \frac{1}{T-t}\textstyle\sum_{h=t+1}^T V_{nh})'P_{n,j}(V_{nt} - \frac{1}{T-t}\sum_{h=t+1}^T V_{nh}) \times Q_{ns}'(-\frac{1}{T-s}\sum_{h=t}^T V_{nh})\right] \\
= \; & c_{Tt}^2 c_{Ts}\mu_3 E Q_{ns}' vec_D(P_{n,j})(-\frac{1}{T-s} - \frac{T-t}{(T-t)^2(T-s)}) = \mu_3 E Q_{ns}' vec_D(P_{n,j}) \cdot (-c_{Ts}\frac{1}{T-s}).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
& E[(\textstyle\sum_{t=1}^{T-1} V_{nt}^{*\prime}P_{n,j}V_{nt}^*)(\sum_{t=1}^{T-1} Q_{nt}'V_{nt}^*)] \\
= \; & \textstyle\sum_{t=1}^{T-1} E\left[(V_{nt}^{*\prime}P_{n,j}V_{nt}^*)(Q_{nt}'V_{nt}^*)\right] + \sum_{s=1}^{T-2}\sum_{t>s}^{T-1} E\left[(V_{nt}^{*\prime}P_{n,j}V_{nt}^*)(Q_{ns}'V_{ns}^*)\right] \\
= \; & \textstyle\sum_{t=1}^{T-1} \mu_3 E Q_{nt}' vec_D(P_{n,j}) \cdot c_{Tt}(1 - \frac{1}{T-t}) - \sum_{s=1}^{T-2} \mu_3 E Q_{ns}' vec_D(P_{n,j}) \cdot c_{Ts}(1 - \frac{1}{T-s}) = 0.
\end{aligned}
$$

Therefore, $\text{Cov}(\mathbf{V}_{n,T-1}^{*\prime}\mathbf{P}_{n,T-1,j}\mathbf{V}_{n,T-1}^*, \mathbf{Q}_{n,T-1}'\mathbf{V}_{n,T-1}^*(\theta)) = 0$. ■

**Proof for Lemma 3:** The objective is sum of three terms: (i) $\frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^{T-1}\left\{c_{Tt}C_{nt}'V_{nt} + V_{nt}'B_nV_{nt} - \sigma_0^2 trB_n\right\}$, (ii) $-\frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^{T-1}\left\{\frac{1}{(T-t+1)^{1/2}(T-t)^{1/2}}C_{nt}'\sum_{h=t+1}^T V_{nh}\right\}$ and (iii) $-\frac{T}{\sqrt{n(T-1)}}\{\bar{V}_{nT}'B_n\bar{V}_{nT} - \frac{1}{T}\sigma_0^2 trB_n\}$. The first term will obey CLT by using Theorem 13 in Yu et al. (2008), where $\lim_{n\to\infty}\frac{1}{n(T-1)}\sum_{t=1}^{T-1}c_{Tt}^2 EC_{nt}'C_{nt} = \lim_{n\to\infty}\frac{1}{n(T-1)}\sum_{t=1}^{T-1}EC_{nt}'C_{nt}$ because $c_{Tt}^2 = \frac{T-t}{T-t+1}$. For the second term, its expectation is zero and its variance is $\frac{1}{n(T-1)}\sum_{t=1}^{T-1}\left[\frac{1}{(T-t+1)}EC_{nt}'C_{nt}\right] < \frac{c}{(T-1)}\sum_{t=1}^{T}\left[\frac{1}{(T-t+1)}\right] = O\left(\frac{\ln T}{T}\right)$ for a finite constant $c$ as $\frac{1}{n}EC_{nt}'C_{nt}$ is bounded uniformly in $t$. The third term will be $O_p\left(\frac{1}{\sqrt{nT}}\right)$ by Lemma 9 in Yu et al. (2008). Hence, for large $T$, as the last two terms will vanish, the CLT follows directly from the first term.

For the case of a finite $T$, the second and third terms would not vanish, but can be combined with the first term into a linear and quadratic system in terms of $(V_{n1}', ..., V_{nT}')'$. The asymptotic will rely on $n \to \infty$. The linear term would involve predetermined variables in its coefficients (instead of constants as in

Kelejian and Prucha (2001)). However, as the proof of the CLT in Prucha and Kelejian (2001) is based on the martingale CLT, it can be extended, similarly to Yu et al. (2008), to cover the predetermined variables situation without additional complication. ■

**Proof for Lemma 4:** This is extended from Liu et al. (2006). ■

**Proof for Lemma 5:** With $Y_{nt}^w = Y_{nt} - (I_n - A_n)^{-1}S_n^{-1}\mathbf{c}_{n0}$, from (1), we have

$$Y_{nt}^w = \lambda_0 W_n Y_{nt}^w + \gamma_0 Y_{n,t-1}^w + \rho_0 W_n Y_{n,t-1}^w + X_{nt}\beta_0 + V_{nt}, \quad t = 1, 2, ..., T. \tag{35}$$

Also, with the Helmert transformation to eliminate individual effects, we have

$$Y_{n,t-1}^{(*,-1)} = c_{Tt}(Y_{n,t-1} - \frac{1}{T-t}\sum_{s=t}^{T-1} Y_{ns}) = c_{Tt}(Y_{n,t-1}^w - \frac{1}{T-t}\sum_{s=t}^{T-1} Y_{ns}^w). \tag{36}$$

We expand $Y_{n,t+h}^w$ for $h \geq 0$ so that $Y_{n,t+h}^w = A_n^{h+1}Y_{n,t-1}^w + S_n^{-1}\sum_{j=0}^h A_n^j X_{n,t+h-j}\beta_0 + S_n^{-1}\sum_{j=0}^h A_n^j V_{n,t+h-j}$. Therefore, we have $\sum_{s=t}^{T-1} Y_{ns}^w = \sum_{h=1}^{T-t} A_n^h Y_{n,t-1}^w + S_n^{-1}\sum_{r=t}^{T-1}(\sum_{h=0}^{T-r-1} A_n^h)(X_{nr}\beta_0 + V_{nr})$. Thus, we can rewrite (36) as $Y_{n,t-1}^{(*,-1)} = c_{Tt}\left[(I_n - \frac{A_n\Phi_{T-t}}{T-t})Y_{n,t-1}^w - \tilde{X}_{n,tT}\beta_0 - \tilde{V}_{n,tT}\right]$. With $E(Y_{n,t-1}^{(*,-1)}|\mathcal{I}_{t-1}) = \Psi_t Y_{n,t-1}^w - c_{Tt}\tilde{X}_{n,tT}\beta_0$, the result follows. ■

**Proof for Lemma 6:** From (11), we can decompose $\mathbb{H}_{nt}$ into $\mathbb{H}_{nt} = \mathbb{H}_{nt}^X + \mathbb{H}_{nt}^V$, where $\mathbb{H}_{nt}^X$ is the deterministic part and $\mathbb{H}_{nt}^V$ is the stochastic part which has zero mean. With $Y_{ns} - A_n Y_{n,s-1} = S_n^{-1}(X_{ns}\beta_0 + \mathbf{c}_{n0} + V_{ns})$ and $Y_{n0} = S_n^{-1}\sum_{h=0}^{h^*} A_n^h X_{nh} + S_n^{-1}\sum_{h=0}^{h^*} A_n^h V_{nh}$, we have, for $t \geq 2$, $\mathbb{H}_{nt}^X = \Psi_t S_n^{-1}\sum_{h=0}^{t-1+h^*} A_n^h X_{n,t-1+h} - c_{Tt}\tilde{X}_{n,tT}\beta_0 - \Psi_t(I_n - A_n)^{-1}S_n^{-1}\mathbf{c}_{n0}$ and $\mathbb{H}_{nt}^V = \Psi_t S_n^{-1}\sum_{h=0}^{t-1+h^*} A_n^h V_{n,t-1+h} - \Psi_t(I_n - A_n)^{-1}S_n^{-1}\frac{1}{t-1}\sum_{s=1}^{t-1} V_{ns}$. For $t = 1$, we have $\mathbb{H}_{n1}^X = \Psi_1 S_n^{-1}\sum_{h=0}^{h^*} A_n^h X_{nh} - c_{T1}\tilde{X}_{n,1T}\beta_0$ and $\mathbb{H}_{n1}^V = \Psi_1 S_n^{-1}\sum_{h=0}^{h^*} A_n^h V_{nh}$. Thus, elements of $\mathbb{H}_{nt}^X$ for $t = 1, ..., T-1$ are $O(1)$, and $\mathbb{H}_{nt}^V$'s are moving averages of past disturbances. By Lemma 7 in Yu et al. (2008), $E(\frac{1}{n(T-1)}\sum_{t=1}^{T-1}\mathbb{H}_{nt}'^V B_n \mathbb{H}_{nt}^V) = O(1)$ and $\text{Var}(\frac{1}{n(T-1)}\sum_{t=1}^{T-1}\mathbb{H}_{nt}'^V B_n\mathbb{H}_{nt}^V) = O(\frac{1}{nT})$. Also, $E\mathbb{H}_{nt}'^X B_n \mathbb{H}_{nt}^V = 0$ and $\text{Var}(\frac{1}{n(T-1)}\sum_{t=1}^{T-1}\mathbb{H}_{nt}'^X B_n\mathbb{H}_{nt}^V) = O(\frac{1}{nT})$. Thus, $\frac{1}{n(T-1)}\sum_{t=1}^{T-1}\mathbb{H}_{nt}' B_n \mathbb{H}_{nt} = E(\frac{1}{n(T-1)}\sum_{t=1}^{T-1}\mathbb{H}_{nt}' B_n\mathbb{H}_{nt}) + O_p(\frac{1}{\sqrt{nT}})$ with $E(\frac{1}{n(T-1)}\sum_{t=1}^{T-1}\mathbb{H}_{nt}' B_n \mathbb{H}_{nt}) = O(1)$.

Defining $\Xi_n = (I_n - A_n)^{-1}S_n$. By Lemma 2 in Yu et al. (2008), $E(\mathbb{W}_{nt}' B_n \mathbb{W}_{nt}) = \sigma_0^2 \frac{1}{t-1}\Xi_n' \Psi_t' B_n \Psi_t \Xi_n$ for $t \geq 2$. For $t = 1$, $\mathbb{W}_{n1} = -\Psi_1 \Xi_n \mathbf{c}_{n0}$ so that $\frac{1}{n}\mathbb{W}_{n1}' B_n \mathbb{W}_{n1} = O(1)$ as elements of $\mathbf{c}_{n0}$ are bounded for all $n$. Thus, for some finite constant $c$, $E\frac{1}{n(T-1)}\sum_{t=1}^{T-1}\mathbb{W}_{nt}' B_n \mathbb{W}_{nt} \leq \sigma_0^2 \frac{c}{(T-1)}\sum_{t=2}^{T-1}(\frac{1}{t-1}) + O\left(\frac{1}{T}\right) = O\left(\frac{\ln T}{T}\right) \to 0$. Also, $\text{Var}(\frac{1}{n(T-1)}\sum_{t=1}^{T-1}\mathbb{W}_{nt}' B_n \mathbb{W}_{nt}) = \frac{1}{n^2(T-1)^2}\sum_{t=2}^{T-1}\sum_{s=2}^{T-1}\text{Cov}(\mathbb{W}_{nt}' B_n \mathbb{W}_{nt}, \mathbb{W}_{ns}' B_n \mathbb{W}_{ns})$ because $\mathbb{W}_{n1}$ is nonstochastic. By Lemma 4 in Yu et al. (2008), for $t \geq s$, we have

$$\begin{aligned}
& \text{Cov}(\mathbb{W}_{nt}' B_n \mathbb{W}_{nt}, \mathbb{W}_{ns}' B_n \mathbb{W}_{ns}) \\
= \ & (\mu_4 - 3\sigma_0^4) \cdot (s-1) \cdot vec_D' \left[ \frac{1}{(t-1)^2} (\Psi_t \Xi_n)' B_n \Psi_t \Xi_n \right] \cdot vec_D \left[ \frac{1}{(s-1)^2} (\Psi_s \Xi_n)' B_n (\Psi_s \Xi_n) \right] \\
& + 2\sigma_0^4 \cdot (s-1)^2 \cdot tr \left[ \frac{1}{(t-1)(s-1)} (\Psi_t \Xi_n) B_n (\Psi_s \Xi_n)' \cdot \frac{1}{(t-1)(s-1)} (\Psi_s \Xi_n) B_n' (\Psi_t \Xi_n)' \right] \\
= \ & O\left( \frac{n}{(t-1)^2 (s-1)} \right) + O\left( \frac{n}{(t-1)^2} \right) = O\left( \frac{n}{(t-1)^2} \right).
\end{aligned}$$

Thus, $\text{Var}(\frac{1}{n(T-1)} \sum_{t=1}^{T-1} \mathbb{W}_{nt}' B_n \mathbb{W}_{nt}) \leq \frac{c}{n(T-1)^2} \sum_{t=2}^{T-1} \sum_{s=2}^{T-1} \frac{1}{(t-1)^2} = O\left(\frac{1}{nT}\right)$. Hence, $\frac{1}{n(T-1)} \sum_{t=1}^{T-1} \mathbb{W}_{nt}' B_n \mathbb{W}_{nt} = o_p(1)$. Similarly, $\frac{1}{n(T-1)} \sum_{t=1}^{T-1} (\mathbb{W}_{nt} + \eta_{nt})' B_n (\mathbb{W}_{nt} + \eta_{nt}) = o_p(1)$.

For $\frac{1}{n(T-1)} \sum_{t=1}^{T-1} \mathbb{H}_{nt}' B_n \mathbb{W}_{nt}$, it has two components $\frac{1}{n(T-1)} \sum_{t=1}^{T-1} \mathbb{H}_{nt}^{X\prime} B_n \mathbb{W}_{nt}$ and $\frac{1}{n(T-1)} \sum_{t=1}^{T-1} \mathbb{H}_{nt}^{V\prime} B_n \mathbb{W}_{nt}$. For the first part, $E \frac{1}{n(T-1)} \sum_{t=1}^{T-1} \mathbb{H}_{nt}^{X\prime} B_n \mathbb{W}_{nt} = \frac{1}{n(T-1)} \mathbb{H}_{n1}^{X\prime} B_n \mathbb{W}_{n1} = O\left(\frac{1}{T}\right)$ and $\text{Var}(\frac{1}{n(T-1)} \sum_{t=1}^{T-1} \mathbb{H}_{nt}^{X\prime} B_n \mathbb{W}_{nt}) = \frac{1}{n^2 (T-1)^2} \sum_{t=2}^{T-1} \sum_{s=2}^{T-1} \mathbb{H}_{nt}^{X\prime} B_n (E \mathbb{W}_{nt} \mathbb{W}_{ns}') B_n' \mathbb{H}_{ns}^X$. As elements of $\mathbb{H}_{nt}^X$ are bounded uniformly in $n$ and $t$, and $E \mathbb{W}_{nt}' \mathbb{W}_{ns} = O\left(\frac{1}{t-1}\right)$ for $t \geq s$, we have $\text{Var}(\sum_{t=1}^{T-1} \mathbb{H}_{nt}^{X\prime} B_n \mathbb{W}_{nt}) = O(nT \ln T)$. Hence, $\frac{1}{n(T-1)} \sum_{t=1}^{T-1} \mathbb{H}_{nt}^{X\prime} B_n \mathbb{W}_{nt} = o_p(1)$. For the second part, by using Lemma 4 in Yu et al. (2008), $\frac{1}{n(T-1)} \sum_{t=1}^{T-1} \mathbb{H}_{nt}^{V\prime} B_n \mathbb{W}_{nt} = o_p(1)$. Thus, $\frac{1}{n(T-1)} \sum_{t=1}^{T-1} \mathbb{H}_{nt}' B_n \mathbb{W}_{nt} = o_p(1)$. Similarly, we have $\frac{1}{n(T-1)} \sum_{t=1}^{T-1} \mathbb{H}_{nt}' B_n (\mathbb{W}_{nt} + \eta_{nt}) = o_p(1)$. ■

**Proof for Lemma 7:**

For (i), because $B_{1n} B_{1n}'$ is non-negative definite, $B_{1n} B_{1n}' = \Gamma_n \Lambda_n \Gamma_n'$ where $\Gamma_n$ is an orthonormal matrix and $\Lambda_n$ is the eigenvalue matrix. It follows that $M_{nt} B_{1n} B_{1n}' M_{nt} \leq \bar{\lambda}_n M_{nt}$, where $\bar{\lambda}_n$ is the largest eigenvalue. By the spectral radius theorem, $tr(M_{nt} B_{1n} B_{1n}' M_{nt}) \leq ||B_{1n} B_{1n}'|| tr(M_{nt}) \leq c K_t$ where $|| \cdot ||$ denotes either the row or column sum norm, and $c$ is some constant such that $||B_{1n} B_{1n}'|| \leq c$ because $B_{1n}$ is UB.

For (ii), as $tr(B_{1n} M_{nt} B_{2n}) = tr(B_{2n} B_{1n} M_{nt})$, it is sufficient to show the case for $tr(B_{1n} M_{nt})$. By the Cauchy-Schwarz inequality, $|tr(B_{1n} M_{nt})| \leq tr^{\frac{1}{2}}(B_{1n} M_{nt} B_{1n}') tr^{\frac{1}{2}}(M_{nt}) \leq c K_t$, because $tr(B_{1n} M_{nt} B_{1n}') = tr(M_{nt} B_{1n}' B_{1n} M_{nt}) \leq c^2 K_t$ for some $c$ by (1).

For (iii), by the Cauchy-Schwarz inequality, $|tr(M_{nt} B_{1n} M_{ns} B_{2n})| \leq [tr(B_{1n}' M_{nt} B_{1n})]^{1/2} [tr(B_{2n}' M_{ns} B_{2n})]^{1/2}$. As $tr(B_{1n}' M_{nt} B_{1n}) = tr(M_{nt} B_{1n} B_{1n}' M_{nt}) \leq ||B_{1n} B_{1n}'|| tr(M_{nt}) \leq c K_t$, the lemma follows. ■

**Proof For Lemma 8:** Because $E(V_{nt}^{*\prime} B_n M_{nt} V_{nt}^*) = E[E(V_{nt}^{*\prime} B_n M_{nt} V_{nt}^* | \mathcal{I}_{t-1})] = \sigma_0^2 E[tr(B_n M_{nt})] = O(K_t)$, $E(\sum_{t=1}^{T-1} V_{nt}^{*\prime} B_n M_{nt} V_{nt}^*) = O(\sum_{t=1}^{T-1} K_t)$. Also, from Lemma 1 (v),

$$\begin{aligned}
E[\text{Var}(V_{nt}^{*\prime} B_n M_{nt} V_{nt}^*) | \mathcal{I}_{t-1}] \ = \ & (\mu_4 - 3\sigma_0^4) c_{Tt}^4 \left( 1 + \frac{1}{(T-t)^3} \right) E[vec_D'(B_n M_{nt}) vec_D(B_n M_{nt})] \\
& + \sigma_0^4 E[tr(B_n M_{nt} B_n') + tr(B_n M_{nt} B_n M_{nt})].
\end{aligned}$$

As $vec_D'(B_nM_{nt})vec_D(B_nM_{nt}) \leq tr(M_{nt}B_n'B_nM_{nt})$, by Lemma 7, $E[\text{Var}(V_{nt}^{*\prime}B_nM_{nt}V_{nt}^*)|\mathcal{I}_{t-1}] = O(K_t)$. Because

$$\begin{aligned}
& \text{Var}(V_{nt}^{*\prime}B_nM_{nt}V_{nt}^* - \sigma_0^2 tr(B_nM_{nt})) \\
= \; & E\{\text{Var}(V_{nt}^{*\prime}B_nM_{nt}V_{nt}^* - \sigma_0^2 tr(B_nM_{nt})|\mathcal{I}_{t-1})\} + \text{Var}\{E(V_{nt}^{*\prime}B_nM_{nt}V_{nt}^* - \sigma_0^2 tr(B_nM_{nt})|\mathcal{I}_{t-1})\} \\
= \; & E\{\text{Var}(V_{nt}^{*\prime}B_nM_{nt}V_{nt}^* - \sigma_0^2 tr(B_nM_{nt})|\mathcal{I}_{t-1})\} = E\{\text{Var}(V_{nt}^{*\prime}B_nM_{nt}V_{nt}^*|\mathcal{I}_{t-1})\} = O(K_t),
\end{aligned}$$

it follows that

$$\begin{aligned}
& \text{Var}(\textstyle\sum_{t=1}^{T-1}(V_{nt}^{*\prime}B_nM_{nt}V_{nt}^* - \sigma_0^2 tr(B_nM_{nt}))) = E(\sum_{t=1}^{T-1}(V_{nt}^{*\prime}B_nM_{nt}V_{nt}^* - \sigma_0^2 tr(B_nM_{nt})))^2 \\
\leq \; & \textstyle\sum_{t=1}^{T-1}\sum_{s=1}^{T-1} \text{Var}^{1/2}(V_{nt}^{*\prime}B_nM_{nt}V_{nt}^* - \sigma_0^2 tr(B_nM_{nt}))\text{Var}^{1/2}(V_{ns}^{*\prime}B_nM_{ns}V_{ns}^* - \sigma_0^2 tr(B_nM_{ns})) \\
= \; & O((\textstyle\sum_{t=1}^{T-1}\sqrt{K_t})^2). \blacksquare
\end{aligned}$$

**Proof for Lemma 9:** As $\eta_{nt} = -c_{Tt}\tilde{V}_{n,tT}$, we have $\sum_{t=1}^{T-1}\eta_{nt}'B_n'M_{nt}V_{nt}^* = -\sum_{t=1}^{T-1}U_{Tt}$ where $U_{Tt} = (V_{nT}',...,V_{nt}')A_{nT,t}(V_{nT}',...,V_{nt}')'$ and $A_{nT,t} = \frac{1}{T+1-t}(\mathbf{0},\Phi_1,...,\Phi_{T-t})'S_n'^{-1}B_n'M_{nt}(-\frac{1}{T-t}I_n,...,-\frac{1}{T-t}I_n,I_n)$. Note that $tr(A_{nT,t}) = \frac{1}{T+1-t}tr(S_n'^{-1}B_n'M_{nt}C_{nTt}')$ where $C_{nTt} = \Phi_{T-t} - \frac{1}{T-t}\sum_{s=t+1}^{T-1}\Phi_{T-s} = \frac{1}{T-t}\sum_{h=1}^{T-t}hA_n^{h-1}$,

$$tr(A_{nT,t}^2) = \frac{1}{(T+1-t)^2}tr(S_n'^{-1}B_n'M_{nt}C_{nTt}'S_n'^{-1}B_n'M_{nt}C_{nTt}'),$$

and $tr(A_{nT,t}A_{nT,t}') = \frac{1}{(T+1-t)(T-t)}tr(S_n'^{-1}B_n'M_{nt}B_nS_n^{-1}\sum_{s=1}^{T-t}\Phi_s\Phi_s')$. As $\sum_{h=1}^{T-t}hA_n^{h-1}$ is UB implied by Assumption 7, elements of $C_{Tt}$ are $O(\frac{1}{T-t})$. Thus, by Lemma 7, $tr(A_{nT,t}) = O\left(\frac{K_t}{(T+1-t)(T-t)}\right)$, $tr(A_{nT,t}^2) = O\left(\frac{K_t}{(T+1-t)^2(T-t)^2}\right)$, and $tr(A_{nT,t}A_{nT,t}') = O\left(\frac{K_t}{T+1-t}\right)$. Hence, $E(U_{Tt}) = \frac{\sigma_0^2}{T+1-t}E[tr(M_{nt}C_{nTt}'S_n'^{-1}B_n')]$,which implies $|E(U_{Tt})| \leq \frac{cK_t}{(T+1-t)(T-t)}$ and $|E\sum_{t=1}^{T-1}U_{Tt}| \leq \left|\sum_{t=1}^{T-1}\frac{cK_t}{(T+1-t)(T-t)}\right|$. For the variance, we have

$$\begin{aligned}
\text{Var}(U_{Tt} - \sigma_0^2 tr(A_{nT,t})) &= E\{\text{Var}(U_{Tt} - \sigma_0^2 tr(A_{nT,t})|\mathcal{I}_{t-1})\} + \text{Var}\{E(U_{Tt} - \sigma_0^2 tr(A_{nT,t})|\mathcal{I}_{t-1})\} \\
&= E\{\text{Var}(U_{Tt} - \sigma_0^2 tr(A_{nT,t})|\mathcal{I}_{t-1})\} = E\{\text{Var}(U_{Tt}|\mathcal{I}_{t-1})\} \\
&= (\mu_4 - 3\sigma_0^4)E[vec_D'(A_{nT,t})vec_D(A_{nT,t})] + \sigma_0^4[E(tr(A_{nT,t}^2)) + E(tr(A_{nT,t}A_{nT,t}'))] \\
&= O(\frac{K_t}{T+1-t}).
\end{aligned}$$

Thus, $\text{Var}(\sum_{t=1}^{T-1}(U_{Tt} - \sigma_0^2 tr(A_{nT,t}))) \leq \sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\text{Var}^{1/2}(U_{Tt} - \sigma_0^2 tr(A_{nT,t}))\text{Var}^{1/2}(U_{Ts} - \sigma_0^2 tr(A_{nT,s})) = O((\sum_{t=1}^{T-1}\sqrt{\frac{K_t}{T+1-t}})^2)$.

Similarly, $\sum_{t=1}^{T-1}\eta_{nt}'B_{n1}M_{nt}B_{n2}\eta_{nt} = \sum_{t=1}^{T-1}W_{Tt}$ where $W_{Tt} = (V_{nT}',...,V_{nt}')B_{nT,t}(V_{nT}',...,V_{nt}')'$ with $B_{nT,t} = \frac{1}{(T-t+1)(T-t)}(\mathbf{0},\Phi_1,...,\Phi_{T-t})'S_n'^{-1}B_{n1}'M_{nt}B_{n2}S_n^{-1}(\mathbf{0},\Phi_1,...,\Phi_{T-t})$. As $tr(B_{nT,t}) = O\left(\frac{K_t}{T-t+1}\right)$, $tr(B_{nT,t}^2) = O\left(\frac{K_t}{(T-t+1)^2}\right)$ and $tr(B_{nT,t}B_{nT,t}') = O\left(\frac{K_t}{(T-t+1)^2}\right)$, we have $E(\sum_{t=1}^{T-1}\eta_{nt}'B_{n1}M_{nt}B_{n2}\eta_{nt}) = \sigma_0^2E[\sum_{t=1}^{T-1}tr(B_{nT,t})] = O\left(\sum_{t=1}^{T-1}\frac{K_t}{T-t+1}\right)$ and $\text{Var}(\sum_{t=1}^{T-1}W_{Tt} - \sigma_0^2 tr(B_{nT,t})) = O_p\left(\left(\sum_{t=1}^{T-1}\frac{\sqrt{K_t}}{T-t+1}\right)^2\right)$. $\blacksquare$

**Proof for Lemma 10:** From (17), it is sufficient to show that $\frac{1}{n(T-1)}\sum_{t=1}^{T-1}(E(Y_{n,t-1}^{(*,-1)}|\mathcal{I}_{t-1}) - H_{nt}\cdot\pi_t)'(E(Y_{n,t-1}^{(*,-1)}|\mathcal{I}_{t-1}) - H_{nt}\cdot\pi_t) \xrightarrow{p} 0$ as $n\to\infty$ and $p_n\to\infty$ for some vector $\pi_t$. As $E(Y_{n,t-1}^{(*,-1)}|\mathcal{I}_{t-1}) = \mathbb{H}_{nt} + \mathbb{W}_{nt}$ from (32) and $\frac{1}{n(T-1)}\sum_{t=1}^{T-1}\mathbb{W}_{nt}'\mathbb{W}_{nt} \xrightarrow{p} 0$ as $n\to\infty$ from Lemma 6, we need to show $\frac{1}{n(T-1)}\sum_{t=1}^{T-1}(\mathbb{H}_{nt} - H_{nt}\cdot\pi_t)'(\mathbb{H}_{nt} - H_{nt}\cdot\pi_t) \xrightarrow{p} 0$ as $n\to\infty$ and $p_n\to\infty$, under which $\frac{1}{n(T-1)}\sum_{t=1}^{T-1}(\mathbb{H}_{nt} - H_{nt}\cdot\pi_t)'\mathbb{W}_{nt} \xrightarrow{p} 0$ by using the Cauchy-Schwarz inequality.

We have $\mathbb{H}_{n1} = \Psi_1 Y_{n0} - c_{T1}\tilde{X}_{1T}\beta_0$ and

$$\begin{aligned}\mathbb{H}_{nt} \;=\;& \frac{1}{t-1}\Psi_t(I_n - A_n)^{-1}A_nY_{n0} - \frac{1}{t-1}\Psi_t\textstyle\sum_{s=2}^{t-1}Y_{n,s-1} + \Psi_t(I_n - \frac{1}{t-1}(I_n - A_n)^{-1})Y_{n,t-1}\\ &+\Psi_t(I_n - A_n)^{-1}S_n^{-1}\frac{1}{t-1}\textstyle\sum_{s=1}^{t-1}X_{ns}\beta_0 - c_{Tt}\tilde{X}_{n,tT}\beta_0,\end{aligned}$$

for $t>1$ from (11). Without loss of generality, consider $\frac{1}{t-1}\Psi_t(I_n - A_n)^{-1}A_nY_{n0}$ with $t\neq 1$. As $A_n = S_n^{-1}(\gamma_0 I_n + \rho_0 W_n) = \gamma_0 I_n + (\gamma_0\lambda_0 + \rho_0)\sum_{j=1}^{\infty}\lambda_0^{j-1}W_n^j$, $A_n^j$ can be written as spatial power series, denoted as $\sum_{h=0}^{\infty}a_j^{(h)}W_n^h$ where $a_j^{(h)}$ is the scalar coefficient for the spatial expansion of $A_n^j$. As $\Phi_j = \sum_{h=0}^{j-1}A_n^h$, $A_n\Phi_{T-t}$ can be written as $\sum_{h=0}^{\infty}(\sum_{j=1}^{T-t}a_j^{(h)})W_n^h$ and $(I_n - A_n)^{-1}A_n$ can be written as $\sum_{h=0}^{\infty}(\sum_{j=1}^{\infty}a_j^{(h)})W_n^h$. Hence, as $\Psi_t = c_{Tt}(I_n - \frac{A_n\Phi_{T-t}}{T-t})$, $\frac{1}{t-1}\Psi_t(I_n - A_n)^{-1}A_n$ can be written as $\frac{c_{Tt}}{t-1}\sum_{h=0}^{\infty}b^{(h)}W_n^h$ where $b^{(0)} = (1 - \frac{\sum_{j=1}^{T-t}a_j^{(0)}}{T-t})(\sum_{j=1}^{\infty}a_j^{(0)})$ and $b^{(h)} = (1 - \frac{\sum_{j=1}^{T-t}a_j^{(0)}}{T-t})(\sum_{j=1}^{\infty}a_j^{(h)}) - \sum_{l=1}^{h}(\frac{\sum_{j=1}^{T-t}a_j^{(l)}}{T-t})(\sum_{j=1}^{\infty}a_j^{(h-l)})$ for $h\geq 1$. Because $\|\lambda_0 W_n\|_\infty < 1$ by Assumption 4, $\|\lambda_0 W_n\|_\infty^{p_n}$ decreases to zero in an exponential rate. By choosing $p_n = \ln n$, we have $\frac{1}{t-1}\Psi_t(I_n - A_n)^{-1}A_nY_{n0} = \frac{c_{Tt}}{t-1}(\sum_{h=0}^{p_n}b^{(h)}W_n^h + R_{n1})Y_{n0}$ where elements of $R_{n1}$ is $O(\frac{1}{n^c})$ for some $c>0$ and $n^cR_{n1}$ is UB. Hence, for $H_{nt}$ in (16), we can find $\pi_t^{(1)}$ such that $\frac{1}{t-1}\Psi_t(I_n - A_n)^{-1}A_nY_{n0} = H_{nt}\cdot\pi_t^{(1)} + \frac{c_{Tt}}{t-1}R_{n1}\cdot Y_{n0}$.

Similarly, $\frac{1}{t-1}\Psi_t\sum_{s=2}^{t-1}Y_{n,s-1} = H_{nt}\cdot\pi_t^{(2)} + \frac{c_{Tt}}{t-1}R_{n2}\sum_{s=2}^{t-1}Y_{n,s-1}$ for some $\pi_t^{(2)}$ and $\Psi_t(I_n - \frac{1}{t-1}(I_n - A_n)^{-1})Y_{n,t-1} = H_{nt}\cdot\pi_t^{(3)} + c_{Tt}R_{n3}\cdot Y_{n,t-1}$ for some $\pi_t^{(3)}$. For the fourth and fifth components of $\mathbb{H}_{nt}$, they are the linear combinations of $X_{n1},...,X_{n,T-1}$ and $X_{nT}$. With spatial power series expansions, we can similarly obtain $\Psi_t(I_n - A_n)^{-1}S_n^{-1}\frac{1}{t-1}\sum_{s=1}^{t-1}X_{ns}\beta_0 - c_{Tt}\tilde{X}_{n,tT}\beta_0 = H_{nt}\cdot\pi_t^{(4)} + c_{Tt}R_{n4}$ for some $\pi_t^{(4)}$.

Thus, we have $\frac{1}{n(T-1)}\sum_{t=1}^{T-1}(\mathbb{H}_{nt} - H_{nt}\cdot\pi_t)'(\mathbb{H}_{nt} - H_{nt}\cdot\pi_t) \to 0$ as $n\to\infty$ and $p_n\to\infty$. ■

**Proof for Lemma 11:**

(i) As $M_{nt} = H_{nt}(H_{nt}'H_{nt})^{+}H_{nt}'$, $(I_n - M_{nt})H_{nt} = 0$. Hence, $e_f(K) = \frac{1}{n(T-1)}\sum_{t=1}^{T-1}f_{nt}'(I_n - M_{nt})f_{nt} = \frac{1}{n(T-1)}\sum_{t=1}^{T-1}(f_{nt} - H_{nt}\cdot\boldsymbol{\pi}_t)'(I_n - M_{nt})(f_{nt} - H_{nt}\cdot\boldsymbol{\pi}_t) \leq \frac{1}{n(T-1)}\sum_{t=1}^{T-1}(f_{nt} - H_{nt}\cdot\pi_t)'(f_{nt} - H_{nt}\cdot\pi_t)$. By Lemma 10, $e_f(K)\to 0$.

(ii) As $f_{nt}$ and $M_{nt}$ involve variables up to period $t-1$ and $V_{nt}^*$ involves the error terms at or after period $t$, we have $E\frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^{T-1}f_{nt}'(I_n - M_{nt})V_{nt}^* = 0$. Also, using $E(V_{nt}^*V_{ns}^{*\prime}|\mathcal{I}_{t-1}) = 0$ for $t>s$ from Lemma 1, we have $\mathrm{Var}(\frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^{T-1}f_{nt}'(I_n - M_{nt})V_{nt}^*) = \sigma_0^2E\frac{1}{n(T-1)}\sum_{t=1}^{T-1}f_{nt}'(I_n - M_{nt})f_{nt} = \sigma_0^2E[e_f(K)]$. Hence, $\frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^{T-1}f_{nt}'(I_n - M_{nt})V_{nt}^* = O_p((E\Delta_K)^{1/2})$;

(iii) For $\frac{1}{n(T-1)}\sum_{t=1}^{T-1} f_{nt}' M_{nt} B_n V_{nt}^*$, its mean is zero. For $t > s$, $\text{Cov}(f_{nt}' M_{nt} B_n V_{nt}^*, f_{ns}' M_{ns} B_n V_{ns}^*) = E[f_{nt}' M_{nt} B_n E(V_{nt}^* \cdot V_{ns}^{*\prime} | \mathcal{I}_{t-1}) B_n' M_{ns} f_{ns}] = 0$ from Lemma 1; for $t = s$, $\text{Cov}(f_{nt}' M_{nt} B_n V_{nt}^*, f_{ns}' M_{ns} B_n V_{ns}^*) = \sigma_0^2 E[f_{nt}' M_{nt} B_n B_n' M_{nt} f_{nt}]$. Hence, we have $\text{Var}(\sum_{t=1}^{T-1} f_{nt}' M_{nt} B_n V_{nt}^*) = \sigma_0^2 \sum_{t=1}^{T-1} E[f_{nt}' M_{nt} B_n B_n' M_{nt} f_{nt}] \leq \sigma_0^2 \|B_n B_n'\|_\infty \sum_{t=1}^{T-1} E(f_{nt}' f_{nt})$. As $\frac{1}{n(T-1)}\sum_{t=1}^{T-1} E(f_{nt}' f_{nt}) = O(1)$ from Lemma 6, $\text{Var}(\frac{1}{n(T-1)}\sum_{t=1}^{T-1} f_{nt}' M_{nt} B_n V_{nt}^*) = O\left(\frac{1}{n(T-1)}\right)$. Hence, $\frac{1}{n(T-1)}\sum_{t=1}^{T-1} f_{nt}' M_{nt} B_n V_{nt}^* = O_p\left(\frac{1}{\sqrt{nT}}\right)$. For $\frac{1}{n(T-1)}\sum_{t=1}^{T-1} f_{nt}' M_{nt} B_n \tilde{V}_{n,tT}$, its mean is zero and its variance is $\frac{1}{n^2(T-1)^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1} E[f_{ns}' M_{ns} B_n \tilde{V}_{n,sT} \tilde{V}_{n,tT}' B_n' M_{nt} f_{nt}]$. For $s \leq t$, we have $E(\tilde{V}_{n,sT}\tilde{V}_{n,tT}' | \mathcal{I}_{t-1}) = \frac{\sigma_0^2}{(T-s)} S_n^{-1}[\frac{1}{T-t}\sum_{j=1}^{T-t} \Phi_j \Phi_j'] S_n'^{-1} = O\left(\frac{1}{T-s}\right)$. Thus,

$$\begin{aligned}
&\text{Var}(\textstyle\sum_{t=1}^{T-1} f_{nt}' M_{nt} B_n \tilde{V}_{n,tT}) = \sum_{t=1}^{T-1}\sum_{s=1}^{T-1} E[f_{ns}' M_{ns} B_n E(\tilde{V}_{n,sT}\tilde{V}_{n,tT}' | \mathcal{I}_{\max\{t,s\}-1}) B_n' M_{nt} f_{nt}] \\
\leq\ & c_1 \textstyle\sum_{t=1}^{T-1}\sum_{s=1}^{T-1} \dfrac{1}{(T-\min\{t,s\})}[E(f_{ns}' M_{ns} f_{ns})]^{1/2}[E(f_{nt} M_{nt} f_{nt})]^{1/2} \leq n c_2 \sum_{t=1}^{T-1}\sum_{s=1}^{T-1} \dfrac{1}{(T-\min\{t,s\})},
\end{aligned}$$

as $E(\frac{f_{nt}' f_{nt}}{n}) = O(1)$ uniformly in $t$. Because $\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\frac{1}{(T-\min\{t,s\})} = 2(T-1) - \sum_{s=1}^{T-1}\frac{1}{T-s} = O(T-1)$, $\text{Var}(\frac{1}{n(T-1)}\sum_{t=1}^{T-1} f_{nt}' M_{nt} B_n \tilde{V}_{n,tT}) = O\left(\frac{1}{n(T-1)}\right)$.

(iv) From (18), the results follow as they are linear combinations of $\frac{1}{n(T-1)}\sum_{t=1}^{T-1} V_{nt}^{*\prime}\mathcal{B}_{n1} M_{nt}\mathcal{B}_{n2} V_{nt}^*$, $\frac{1}{n(T-1)}\sum_{t=1}^{T-1}\eta_{nt}'\mathcal{B}_{n1} M_{nt}\mathcal{B}_{n2}\eta_{nt}$ and $\frac{1}{n(T-1)}\sum_{t=1}^{T-1}\eta_{nt}'\mathcal{B}_{n1} M_{nt}\mathcal{B}_{n2} V_{nt}^*$, where $\mathcal{B}_{ni}$'s are UB. Thus, from Lemmas 8 and 9, we have the results. ■

# C Alternative Finite Moments in the Systematic Setting

As derived from Lemma 5, we have $Y_{n,t-1}^{(*,-1)} = \Psi_t Y_{n,t-1}^w - c_{Tt}\tilde{X}_{n,tT}\beta_0 - c_{Tt}\tilde{V}_{n,tT}$ where $Y_{n,t-1}^w = Y_{n,t-1} - (I_n - A_n)^{-1} S_n^{-1}\mathbf{c}_{n0}$. To construct an optimal IV for $Y_{n,t-1}^{(*,-1)}$, the systematic IV approach in the main text has individual effects $\mathbf{c}_{n0}$ estimated by observables till $t-1$. An alternative approach is to estimate $\mathbf{c}_{n0}$ with the whole sample. Denoting $\hat{\mathbf{c}}_{nT} = \frac{1}{T}\sum_{t=1}^{T}(S_n(\hat{\lambda}) Y_{nt} - Z_{nt}\hat{\delta})$ where $\hat{\lambda}$ and $\hat{\delta}$ are $\sqrt{nT}$ consistent initial estimates (which could be obtained from some simple IV procedures). The alternative feasible optimal IV for $Y_{n,t-1}^{(*,-1)}$ can be $\hat{\mathbb{H}}_{nt}^a = \hat{\Psi}_t\left[Y_{n,t-1} - (I_n - \hat{A}_n)^{-1}\hat{S}_n^{-1}\hat{\mathbf{c}}_{nT}\right] - c_{Tt}\hat{\tilde{X}}_{n,tT}\hat{\beta}$ for $t = 1, ..., T$. Thus, the feasible best IV for $Z_{nt}^*$ is $\hat{\mathbb{K}}_{nt}^a \equiv (\hat{\mathbb{H}}_{nt}^a, W_n\hat{\mathbb{H}}_{nt}^a, X_{nt}^*)$ and the best IV for $W_n Y_{nt}^*$ is $\hat{G}_n\hat{\mathbb{K}}_{nt}^a\delta_0$. Hence, an IV matrix for $(W_n Y_{nt}^*, Z_{nt}^*)$ can be

$$\hat{\mathbb{Q}}_{nt}^a = (\hat{G}_n\hat{\mathbb{K}}_{nt}^a\hat{\delta},\ \hat{\mathbb{K}}_{nt}^a). \tag{37}$$

**Theorem 8** *Under Assumptions 1-10, suppose we use the moment conditions in (3) where $Q_{nt}$ takes the special form $\hat{\mathbb{Q}}_{nt}^a$ in (37) and $\hat{\mathbf{P}}_{n,T-1}$ is estimated from (8). Suppose that $\hat{\Sigma}_{nT}^{-1} - \Sigma_{nT}^{-1} = o_p(1)$. Then, the BGMME $\hat{\theta}_{b,nT}^a$ has $\sqrt{n(T-1)}(\hat{\theta}_{b,nT}^a - \theta_0) \xrightarrow{d} N\left(0, \mathbf{\Sigma}_b^{-1}\right)$ as $n \to \infty$ and $T \to \infty$.*

**Proof.** From Theorem 4 in Yu et al. (2008), $(\hat{\mathbf{c}}_{nT}-\mathbf{c}_{n0})=\frac{1}{T}\sum_{t=1}^{T}V_{nt}+O_p\left(\frac{1}{\sqrt{nT}}\right)$. Thus,

$$\begin{aligned}
\hat{\mathbb{H}}_{nt}^{a} &= \hat{\Psi}_t Y_{n,t-1}-\hat{\Psi}_t(I_n-\hat{A}_n)^{-1}\hat{S}_n^{-1}\mathbf{c}_{n0}-c_{Tt}\hat{\hat{X}}_{n,tT}\hat{\beta}-\hat{\Psi}_t(I_n-\hat{A}_n)^{-1}\hat{S}_n^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}V_{nt}+O_p\left(\frac{1}{\sqrt{nT}}\right)\right)\\
&= E(Y_{n,t-1}^{(*,-1)}|\mathcal{I}_{t-1})+O_p\left(\frac{1}{\sqrt{nT}}\right)+\hat{\Psi}_t(I_n-\hat{A}_n)^{-1}\hat{S}_n^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}V_{nt}+O_p\left(\frac{1}{\sqrt{nT}}\right)\right)\\
&= E(Y_{n,t-1}^{(*,-1)}|\mathcal{I}_{t-1})+O_p\left(\frac{1}{\sqrt{T}}\right).
\end{aligned}$$

Therefore, $E(Y_{n,t-1}^{(*,-1)}|\mathcal{I}_{t-1})$ of $Y_{n,t-1}^{(*,-1)}$ can be well approximated by $\hat{\mathbb{H}}_{nt}^{a}$. By using Lemma 6, $\hat{\theta}_{b,nT}^{a}$ has the same asymptotic distribution as $\hat{\theta}_{b,nT}$ in Theorem 2. ■

Theorem 8 holds when $T$ is large, where $\frac{1}{T}\sum_{t=1}^{T}V_{nt}$ will vanish in probability as $T\to\infty$, and $\hat{\mathbb{H}}_{nt}^{a}$ would be asymptotically the best IV. On the other hand, when $T$ is finite, $\frac{1}{T}\sum_{t=1}^{T}V_{nt}$ will not vanish and its presence will cause the correlation of $\hat{\mathbb{H}}_{nt}^{a}$ with the disturbances $V_{nt}^{*}$. Thus, when $T$ is finite, the linear moments from (37) may cause inconsistent estimate.

# D Proofs for Theorems

## D.1 Proof for Theorem 1

We first derive the uniform convergence of $\frac{1}{n(T-1)}a_{nT}g_{nT}(\theta)$. Combined with the identification in Assumption 9, the consistency of GMM estimator $\hat{\theta}_{nT}$ will follow. Let $a_{nT}=(a_{nT}^{(1)},\cdots a_{nT}^{(m)},a_{nT}^{(Q)})$ be a $k_a\times(m+q)$ matrix. Then,

$$\frac{1}{n(T-1)}a_{nT}g_{nT}(\theta)=\frac{1}{n(T-1)}\mathbf{V}_{n,T-1}^{*\prime}(\theta)(\textstyle\sum_{j=1}^{m}a_{nT}^{(j)}\mathbf{P}_{n,T-1,l})\mathbf{V}_{n,T-1}^{*}(\theta)+\frac{1}{n(T-1)}a_{nT}^{(Q)}\mathbf{Q}_{n,T-1}'\mathbf{V}_{n,T-1}^{*}(\theta),$$

where, by expansion, $\mathbf{V}_{n,T-1}^{*}(\theta)=\mathbf{d}_{n,T-1}^{*}(\theta)+(I_{n(T-1)}+(\lambda_0-\lambda)\mathbf{G}_{n,T-1})\mathbf{V}_{n,T-1}^{*}$ with $\mathbf{d}_{n,T-1}^{*}(\theta)=(\lambda_0-\lambda)\mathbf{G}_{n,T-1}\mathbf{Z}_{n,T-1}^{*}\delta_0+\mathbf{Z}_{n,T-1}^{*}(\delta_0-\delta)$.

For $\frac{1}{n(T-1)}a_{nT}^{(Q)}\mathbf{Q}_{n,T-1}'\mathbf{V}_{n,T-1}^{*}(\theta)=\frac{1}{n(T-1)}a_{nT}^{(Q)}\mathbf{Q}_{n,T-1}'\mathbf{d}_{n,T-1}^{*}(\theta)+\frac{1}{n(T-1)}a_{nT}^{(Q)}\mathbf{Q}_{n,T-1}'(I_{n(T-1)}+(\lambda_0-\lambda)\mathbf{G}_{n,T-1})\mathbf{V}_{n,T-1}^{*}$, the second term is $o_p(1)$ uniformly in $\theta\in\Theta$ from Lemma 1 (iv). Because

$$\mathbf{V}_{n,T-1}^{*\prime}(\theta)(\textstyle\sum_{j=1}^{m}a_{nT}^{(j)}\mathbf{P}_{n,T-1,l})\mathbf{V}_{n,T-1}^{*}(\theta)=\mathbf{d}_{n,T-1}^{*\prime}(\theta)(\sum_{j=1}^{m}a_{nT}^{(j)}\mathbf{P}_{n,T-1,l})\mathbf{d}_{n,T-1}^{*}(\theta)+\mathbf{l}_{n,T-1}(\theta)+\mathbf{q}_{n,T-1}(\theta)$$

where $\mathbf{l}_{n,T-1}(\theta)=\mathbf{d}_{n,T-1}^{*\prime}(\theta)(\sum_{j=1}^{m}a_{nT}^{(j)}\mathbf{P}_{n,T-1,l})(\mathbf{V}_{n,T-1}^{*}+(\lambda_0-\lambda)\mathbf{G}_{n,T-1}\mathbf{V}_{n,T-1}^{*})$ and

$$\mathbf{q}_{n,T-1}(\theta)=\left(\mathbf{V}_{n,T-1}^{*}+(\lambda_0-\lambda)\mathbf{G}_{n,T-1}\mathbf{V}_{n,T-1}^{*}\right)'(\textstyle\sum_{j=1}^{m}a_{nT}^{(j)}\mathbf{P}_{n,T-1,l})\left(\mathbf{V}_{n,T-1}^{*}+(\lambda_0-\lambda)\mathbf{G}_{n,T-1}\mathbf{V}_{n,T-1}^{*}\right),$$

it is sufficient to prove that $\frac{1}{n(T-1)}\mathbf{l}_{n,T-1}(\theta)$ and $\frac{1}{n(T-1)}\mathbf{q}_{n,T-1}(\theta)$ converge uniformly to well defined limits. By Lemma 1, as $\frac{1}{n(T-1)}\mathbf{l}_{n,T-1}(\theta)$ will converge to 0 and $\frac{1}{n(T-1)}\mathbf{q}_{n,T-1}(\theta)$ will converge to its mean, the desirable convergence follows. The uniform convergence and identification uniqueness imply the consistency of the estimator.

By the Taylor expansion

$$\sqrt{n(T-1)}(\hat{\theta}_{nT}-\theta_0) = -\left[\frac{\partial g'_{nT}(\hat{\theta}_{nT})/\partial\theta}{n(T-1)} a'_{nT}a_{nT}\frac{\partial g_{nT}(\bar{\theta}_{nT})/\partial\theta'}{n(T-1)}\right]^{-1}\frac{\partial g'_{nT}(\hat{\theta}_{nT})/\partial\theta}{n(T-1)} a'_{nT}a_{nT}\frac{g_{nT}(\theta_0)}{\sqrt{n(T-1)}},$$

where $\bar{\theta}_{nT}$ lies between $\hat{\theta}_{nT}$ and $\theta_0$ and $\frac{\partial g_{nT}(\theta)}{\partial\theta'} = (\frac{\partial g_{nT}(\theta)}{\partial\lambda}, \frac{\partial g_{nT}(\theta)}{\partial\delta'}) = (-1)\times$
$\begin{pmatrix} (\mathbf{W}_{n,T-1}\mathbf{Y}^*_{n,T-1})'\mathbf{P}^s_{n,T-1,1}\mathbf{V}^*_{n,T-1}(\theta) & \cdots & (\mathbf{W}_{n,T-1}\mathbf{Y}^*_{n,T-1})'\mathbf{P}^s_{n,T-1,m}\mathbf{V}^*_{n,T-1}(\theta) & (\mathbf{W}_{n,T-1}\mathbf{Y}^*_{n,T-1})'\mathbf{Q}_{n,T-1} \\ \mathbf{Z}^{*\prime}_{n,T-1}\mathbf{P}^s_{n,T-1,1}\mathbf{V}^*_{n,T-1}(\theta) & \cdots & \mathbf{Z}^{*\prime}_{n,T-1}\mathbf{P}^s_{n,T-1,m}\mathbf{V}^*_{n,T-1}(\theta) & \mathbf{Z}^{*\prime}_{n,T-1}\mathbf{Q}_{n,T-1}\end{pmatrix}'$.
By Lemma 1 and $\hat{\theta}_{nT}-\theta_0 = o_p(1)$, $\frac{1}{n(T-1)}\frac{\partial g_{nT}(\hat{\theta}_{nT})}{\partial\theta'} = D_{nT}+o_p(1)$. Thus, $\frac{\partial g'_{nT}(\hat{\theta}_{nT})/\partial\theta}{n(T-1)}a'_{nT}a_{nT}\frac{\partial g_{nT}(\bar{\theta}_{nT})/\partial\theta'}{n(T-1)} = D'_{nT}a'_{nT}a_{nT}D_{nT}+o_p(1)$. Also, $\frac{1}{\sqrt{n(T-1)}}a_{nT}g_{nT}(\theta_0) \xrightarrow{d} N(0, \text{plim}_{n\to\infty}a_{nT}\Sigma_{nT}a'_{nT})$ from Lemma 3. Hence,

$$\sqrt{n(T-1)}(\hat{\theta}_{nT}-\theta_0) \xrightarrow{d} N(0, \text{plim}_{n\to\infty}\left(D'_{nT}a'_{nT}a_{nT}D_{nT}\right)^{-1}D'_{nT}a'_{nT}a_{nT}\Sigma_{nT}a'_{nT}a_{nT}D_{nT}\left(D_{nT}a'_{nT}a_{nT}D_{nT}\right)^{-1}).$$

For the optimum GMM, $\Sigma^{-1}_{nT}$ is used as $a'_{nT}a_{nT}$, and its efficiency relative to the ones with $a_{nT}$ follows from the generalized Cauchy-Schwarz inequality.

When $\Sigma_{nT}$ is replaced by $\hat{\Sigma}_{nT}$ so that $\hat{\Sigma}_{nT} = \Sigma_{nT}+o_p(1)$, we will have the same asymptotic distribution by similar arguments for Proposition 2 in Lee (2007). ■

## D.2 Proof for Theorem 2

For the variance matrix $(D'_{nT}\Sigma^{-1}_{nT}D_{nT})^{-1}$ in (7) of the OGMME, we shall first show that $\mathbf{P}_{n,T-1}$ in (8) is the best quadratic moment matrix, and $\mathbb{Q}_{nt}$ in (12) is the best linear IV matrix when $T$ is large. We then proceed to prove the consistency and asymptotic distribution of the best GMME using estimated $\hat{\mathbf{P}}_{n,T-1}$ from (8) and $\tilde{\mathbb{Q}}_{nt}$ in (13).

For quadratic moments, from Lemma 4, the $\frac{1}{n(T-1)}C_{mn,T}(\frac{\mu_4-3\sigma_0^4}{\sigma_0^4}\omega'_{nm,T}\omega_{nm,T}+\Delta_{mn,T})^{-1}C'_{mn,T}$ in (7) is maximized at $\frac{1}{n(T-1)}vec'(\mathbf{G}^{-s}_{n,T-1})vec(\mathbf{G}^{-s}_{n,T-1})$ by choosing $\mathbf{P}_{n,T-1}$ in (8). From Lemma 1 (iv), when $T$ is large, $\text{plim}_{n\to\infty}\frac{1}{n(T-1)}\mathbf{Q}'_{n,T-1}(\mathbf{G}_{n,T-1}\mathbf{Z}^*_{n,T-1}\delta_0, \mathbf{Z}^*_{n,T-1}) = \text{plim}_{n\to\infty}\frac{1}{n(T-1)}\mathbf{Q}'_{n,T-1}\mathbb{Q}_{n,T-1}$. Thus,

$$\text{plim}_{n\to\infty}\frac{1}{n(T-1)}(\mathbf{G}_{n,T-1}\mathbf{Z}^*_{n,T-1}\delta_0, \mathbf{Z}^*_{n,T-1})'\mathbf{M}_{Q,nT}(\mathbf{G}_{n,T-1}\mathbf{Z}^*_{n,T-1}\delta_0, \mathbf{Z}^*_{n,T-1}) \leq \text{plim}_{n\to\infty}\frac{1}{n(T-1)}\mathbb{Q}'_{n,T-1}\mathbb{Q}_{n,T-1}$$

when $T$ is large. Therefore, the best IV is $\mathbb{Q}_{n,T-1}$. By Lemma 6, $\text{plim}_{n\to\infty}\frac{1}{n(T-1)}(\mathbf{G}_{n,T-1}\mathbf{Z}^*_{n,T-1}\delta_0, \mathbf{Z}^*_{n,T-1})'$ $(\mathbf{G}_{n,T-1}\mathbf{Z}^*_{n,T-1}\delta_0, \mathbf{Z}^*_{n,T-1}) = \text{plim}_{n\to\infty}\frac{1}{n(T-1)}\mathbb{Q}'_{n,T-1}\mathbb{Q}_{n,T-1}$ .

When we use the best $\mathbf{P}_{n,T-1}$ in (8) and $\mathbb{Q}_{nt}$ in (12), the infeasible moment conditions are $g_{nT}(\theta) = \left(\mathbf{V}^{*\prime}_{n,T-1}(\theta)\mathbf{P}_{n,T-1}\mathbf{V}^*_{n,T-1}(\theta), \mathbf{V}^{*\prime}_{n,T-1}(\theta)\mathbb{Q}_{n,T-1}\right)'$ and the identification and uniform convergence of the GMM objective function can be obtained similar to the proof in Appendix D.1. When we use estimated $\tilde{\mathbb{Q}}_{nt}$ and $\hat{\mathbf{P}}_{n,T-1}$, the feasible moment conditions would be $\tilde{g}_{nT}(\theta) = \left(\mathbf{V}^{*\prime}_{n,T-1}(\theta)\hat{\mathbf{P}}_{n,T-1}\mathbf{V}^*_{n,T-1}(\theta), \mathbf{V}^{*\prime}_{n,T-1}(\theta)\tilde{\mathbb{Q}}_{nt}\right)'$.

First, $||A_n(\tilde{\theta})|| - ||A_n|| = o_p(1)$ under $||\tilde{\theta} - \theta_0|| = o_p(1)$. Because $||A_n||_\infty < 1$, $\sum_{h=1}^{\infty}(||A_n^h(\tilde{\theta})||_\infty - ||A_n^h||_\infty) = (1 - ||A_n(\tilde{\theta})||_\infty)^{-1} - (1 - ||A_n||_\infty)^{-1} = o_p(1)$. Hence, $\sum_{h=0}^{\infty} ||A_n^h(\tilde{\theta}) - A_n^h||_\infty = o_p(1)$ and the elements of $\frac{1}{T-t}\sum_{s=t}^{T-1}\sum_{h=1}^{s-1}(A_n^h(\tilde{\theta}) - A_n)X_{n,s-h}$ are $o_p(1)$ uniformly. Therefore, combined with $\left\|S_n^{-1}(\tilde{\lambda}) - S_n^{-1}\right\|_\infty = o_p(1)$ and $\left\|G_n(\tilde{\lambda}) - G_n\right\|_\infty = o_p(1)$, $\tilde{\mathbb{H}}_{nt} - \mathbb{H}_{nt} = \left\|\tilde{\theta} - \theta_0\right\| \cdot B_{\mathbb{H}_{nt}}$ for some $B_{\mathbb{H}_{nt}}$, of which its elements are bounded uniformly in $n$ and $t$. Similarly, $\mathbf{P}_{n,T-1} - \hat{\mathbf{P}}_{n,T-1} = o_p(1) \cdot \mathbf{B}_{\mathbf{P}_{n,T-1}}$ for some $\mathbf{B}_{\mathbf{P}_{n,T-1}}$ which also has a block diagonal pattern similar to $\mathbf{P}_{n,T-1}$ with its diagonal matrices being UB.

Thus, $\frac{1}{n(T-1)}\tilde{g}_{nT}(\theta) = \frac{1}{n(T-1)}\left(\mathbf{V}_{n,T-1}^{*\prime}(\theta)(\hat{\mathbf{P}}_{n,T-1} - \mathbf{P}_{n,T-1})\mathbf{V}_{n,T-1}^{*}(\theta), \mathbf{V}_{n,T-1}^{*\prime}(\theta)(\tilde{\mathbb{Q}}_{n,T-1} - \mathbb{Q}_{n,T-1})\right)' + \frac{1}{n(T-1)}g_{nT}(\theta)$. As $\tilde{\mathbb{H}}_{nt} - \mathbb{H}_{nt} = \left\|\tilde{\theta} - \theta_0\right\| \cdot B_{\mathbb{H}_{nt}}$, the $\frac{1}{n(T-1)}(\tilde{\mathbb{Q}}_{n,T-1} - \mathbb{Q}_{n,T-1})'\mathbf{V}_{n,T-1}^{*}(\theta) = \frac{1}{n(T-1)}(\tilde{\mathbb{Q}}_{n,T-1} - \mathbb{Q}_{n,T-1})'\mathbf{d}_{n,T-1}^{*}(\theta) + \frac{1}{n(T-1)}(\tilde{\mathbb{Q}}_{n,T-1} - \mathbb{Q}_{n,T-1})'(I_{n(T-1)} + (\lambda_0 - \lambda)\mathbf{G}_{n,T-1})\mathbf{V}_{n,T-1}^{*}$ will be $o_p(1)$ uniformly in $\theta$ because $\left\|\tilde{\theta} - \theta_0\right\| = o_p(1)$. Similarly, $\frac{1}{n(T-1)}\mathbf{V}_{n,T-1}^{*\prime}(\theta)(\hat{\mathbf{P}}_{n,T-1} - \mathbf{P}_{n,T-1})\mathbf{V}_{n,T-1}^{*}(\theta)$ is $o_p(1)$ uniformly in $\theta$. Thus, the identification of $\frac{1}{n(T-1)}g_{nT}(\theta)$ implies the identification of $\frac{1}{n(T-1)}\tilde{g}_{nT}(\theta)$ and the uniform convergence of $\frac{1}{n(T-1)}a_{nT}g_{nT}(\theta)$ will imply the uniform convergence of $\frac{1}{n(T-1)}a_{nT}\tilde{g}_{nT}(\theta)$. Hence, the consistency of the estimates using the feasible moments follows.

With $\mathbb{D}_{nT} = -\frac{1}{n(T-1)}\begin{pmatrix} \sigma_0^2 tr(\mathbf{G}_{n,T-1}'\mathbf{P}_{n,T-1}^s) & (\mathbf{G}_{n,T-1}\mathbf{Z}_{n,T-1}^{*}\delta_0)'\mathbb{Q}_{n,T-1} \\ \mathbf{0} & \mathbf{Z}_{n,T-1}^{*\prime}\mathbb{Q}_{n,T-1} \end{pmatrix}'$, we have $\frac{1}{n(T-1)}\frac{\partial \tilde{g}_{nT}(\hat{\theta}_{nT})}{\partial \theta'} = \mathbb{D}_{nT} + o_p(1)$. Also, as $\tilde{\mathbb{H}}_{nt} - \mathbb{H}_{nt} = \left\|\tilde{\theta} - \theta_0\right\| \cdot B_{\mathbb{H}_{nt}}$ and $\mathbf{P}_{n,T-1} - \hat{\mathbf{P}}_{n,T-1} = o_p(1) \cdot \mathbf{B}_{\mathbf{P}_{n,T-1}}$, $\frac{1}{\sqrt{n(T-1)}}(\tilde{g}_{nT}(\theta_0) - g_{nT}(\theta_0)) = \frac{1}{\sqrt{n(T-1)}}(\mathbf{V}_{n,T-1}^{*\prime}(\hat{\mathbf{P}}_{n,T-1} - \mathbf{P}_{n,T-1})\mathbf{V}_{n,T-1}^{*}, \mathbf{V}_{n,T-1}^{*\prime}(\tilde{\mathbb{Q}}_{n,T-1} - \mathbb{Q}_{n,T-1}))' = o_p(1)$. Thus, the GMME obtained from the feasible moments have the same asymptotic distribution as the infeasible ones. Thus, $\sqrt{n(T-1)}(\hat{\theta}_{b,nT} - \theta_0) \xrightarrow{d} N\left(0, \mathbf{\Sigma}_c^{-1}\right)$ with $\mathbf{\Sigma}_c$ in (15). When $T \to \infty$, $\mathbf{\Sigma}_c = \mathbf{\Sigma}_b$. ■

## D.3 Proof for Theorem 3

For (19), denote $\hat{H} = \frac{1}{n(T-1)}\sum_{t=1}^{T-1}(f_{nt} + u_{nt})'M_{nt}(f_{nt} + u_{nt})$ and $\hat{h} = \frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^{T-1}(f_{nt} + u_{nt})'M_{nt}V_{nt}^{*}$. For the 2SLSE, we have $\sqrt{n(T-1)}(\hat{\theta}_{2sl,nT} - \theta_0) = [\hat{H}]^{-1} \times \hat{h}$ where

$$\hat{H} = H + \sum_{i=1}^{3} Z_i^H \quad \text{and} \quad \hat{h} = h + \sum_{i=1}^{2} T_i^h \tag{38}$$

with $H = \frac{1}{n(T-1)}\sum_{t=1}^{T-1} f_{nt}'f_{nt}$, $h = \frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^{T-1} f_{nt}'V_{nt}^{*}$, $Z_1^H = -\frac{1}{n(T-1)}\sum_{t=1}^{T-1} f_{nt}'(I_n - M_{nt})f_{nt} = -e_f(K)$, $Z_2^H = \frac{1}{n(T-1)}\sum_{t=1}^{T-1} f_{nt}'M_{nt}u_{nt} + \frac{1}{n(T-1)}\sum_{t=1}^{T-1} u_{nt}'M_{nt}f_{nt}$, $Z_3^H = \frac{1}{n(T-1)}\sum_{t=1}^{T-1} u_{nt}'M_{nt}u_{nt}$, $T_1^h = -\frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^{T-1} f_{nt}'(I_n - M_{nt})V_{nt}^{*}$ and $T_2^h = \frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^{T-1} u_{nt}'M_{nt}V_{nt}^{*}$.

For the terms in $\hat{H}$, we have $H = O_p(1)$ from Lemma 6. $Z_1^H = O_p(E(\Delta_K)) = o_p(1)$ as $K \to \infty$ from Lemma 11 (i); $Z_2^H = O_p\left(\sqrt{\frac{1}{nT}}\right)$ from Lemma 11 (iii), and $Z_3^H = O_p\left(\frac{\sum_{t=1}^{T-1} K_t}{nT}\right)$ from Lemma 11 (iv). Therefore, $\hat{H} = H + O_p\left(\frac{\sum_{t=1}^{T-1} K_t}{nT}\right) + o_p(1)$.

For the terms in $\hat{h}$, $h$ will be asymptotically normally distributed by Lemma 3 as $\frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^{T-1} f_{nt}'V_{nt}^{*} \xrightarrow{d} N(0, \sigma_0^2 \text{plim}_{n\to\infty}\mathbf{\Sigma}_{nT,22})$. For the residual terms, $T_1^h = O_p(\sqrt{E(\Delta_K)}) = o_p(1)$ from Lemma 11 (ii); $T_2^h$

has two components from (18) which are $T^h_{2,1} = \frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^{T-1}(G_nV^*_{nt}, 0, 0, \mathbf{0}_{n\times k_x})'M_{nt}V^*_{nt}$ and $T^h_{2,2} = \frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^{T-1}(G_n(\eta_{nt}\gamma_0 + W_n\eta_{nt}\rho_0), \eta_{nt}, W_n\eta_{nt}, \mathbf{0}_{n\times k_x})'M_{nt}V^*_{nt}$. By Lemma 8, $T^h_{2,1} = \varphi_1 + (T^h_{2,1} - \varphi_1)$, where $\varphi_1 = O_p\left(\frac{\sum_{t=1}^{T-1}K_t}{\sqrt{n(T-1)}}\right)$ is the conditional mean of $T^h_{2,1}$ and $(T^h_{2,1} - \varphi_1) = O_p\left(\frac{\sum_{t=1}^{T-1}\sqrt{K_t}}{\sqrt{n(T-1)}}\right)$, which is not large than the order $O_p\left(\sqrt{\frac{TK}{n}}\right)$. By Lemma 9, $T^h_{2,2} = \varphi_2 + (T^h_{2,2} - \varphi_2)$ where $\varphi_2 = O_p\left(\frac{1}{\sqrt{nT}}\sum_{t=1}^{T-1}\frac{K_t}{(T+1-t)(T-t)}\right)$ and $(T^h_{2,2} - \varphi_2) = O_p\left(\frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^{T-1}\sqrt{\frac{K_t}{T+1-t}}\right)$. Hence, $\hat{h} = h + \varphi_1 + \varphi_2 + e_1 O_p\left(\frac{\sum_{t=1}^{T-1}\sqrt{K_t}}{\sqrt{n(T-1)}}\right) + O_p\left(\frac{\sum_{t=1}^{T-1}\sqrt{\frac{K_t}{T+1-t}}}{\sqrt{n(T-1)}}\right) + o_p(1)$. We see that $\varphi_2/\varphi_1 \xrightarrow{p} 0$ when $\frac{K}{\sum_{t=1}^{T-1}K_t} \to 0$. Therefore, in $T^h_2$, the spatial endogeneity component $T^h_{2,1}$ dominates $T^h_{2,2}$. This implies that, for the bias of the estimates due to many moments, the dominant term is caused by the spatial endogeneity component.

Combining the expansions of $\hat{H}$ and $\hat{h}$ with $\lim_{n\to\infty} H = \text{plim}_{n\to\infty}\mathbf{\Sigma}_{nT,22}$ from Lemma 6, (20) follows.

Let $\hat{\sigma}^2_{nT} = \frac{1}{n(T-1)}[\mathbf{S}_{n,T-1}(\hat{\lambda}_{2sl,nT})\mathbf{Y}^*_{n,T-1} - \mathbf{Z}^*_{n,T-1}\hat{\delta}_{2sl,nT}]'[\mathbf{S}_{n,T-1}(\hat{\lambda}_{2sl,nT})\mathbf{Y}^*_{n,T-1} - \mathbf{Z}^*_{n,T-1}\hat{\delta}_{2sl,nT}]$. As $\hat{\theta}_{2sl,nT} - \theta_0 = O_p\left(\max\left(\frac{\sum_{t=1}^{T-1}K_t}{n(T-1)}, \frac{1}{\sqrt{n(T-1)}}\right)\right)$ from (20), $\hat{\sigma}^2_{nT} - \sigma^2_0 = O_p\left(\max\left(\frac{\sum_{t=1}^{T-1}K_t}{n(T-1)}, \frac{1}{\sqrt{n(T-1)}}\right)\right)$. With $G_n(\hat{\lambda}_{2sls,nT}) - G_n = G^2_n(\bar{\lambda}_{nT})(\hat{\lambda}_{2sls,nT} - \lambda_0)$ where $G^2_n(\bar{\lambda}_{nT})$ is UB in probability, $\sum_{t=1}^{T-1}[tr(G_nM_{nt})]$ and $\sum_{t=1}^{T-1}[tr(G^2_n(\bar{\lambda}_{nT})M_{nt})]$ of order $O\left(\sum_{t=1}^{T-1}K_t\right)$, we have

$$\begin{aligned}
\hat{\varphi}_1 - \varphi_1 &= \frac{1}{\sqrt{n(T-1)}}\left(\hat{\sigma}^2_{nT}\sum_{t=1}^{T-1}[tr(G_n(\hat{\lambda}_{2sls,nT})M_{nt})] - \sigma^2_0\sum_{t=1}^{T-1}[tr(G_nM_{nt})]\right)e_1 \\
&= O_p\left(\max\left(\frac{1}{\sqrt{n(T-1)}}\left(\frac{\sum_{t=1}^{T-1}K_t}{\sqrt{n(T-1)}}\right)^2, \frac{\sum_{t=1}^{T-1}K_t}{n(T-1)}\right)\right).
\end{aligned}$$

Thus, with $\varphi_2 = O_p\left(\frac{K}{\sqrt{nT}}\right)$, we have

$$\begin{aligned}
&\sqrt{n(T-1)}(\hat{\theta}^1_{2sl,nT} - \theta_0) + O_p\left(\max\left(\frac{1}{\sqrt{n(T-1)}}\left(\frac{\sum_{t=1}^{T-1}K_t}{\sqrt{n(T-1)}}\right)^2, \frac{\sum_{t=1}^{T-1}K_t}{n(T-1)}, \frac{K}{\sqrt{nT}}, \frac{\sum_{t=1}^{T-1}\sqrt{K_t}}{\sqrt{n(T-1)}}\right)\right) \\
&\xrightarrow{d} N(0, \sigma^2_0\text{plim}_{n\to\infty}\mathbf{\Sigma}^{-1}_{nT,22}).
\end{aligned}$$

Under $\frac{\sum_{t=1}^{T-1}K_t}{\sqrt{n(T-1)}} \to c$, $\frac{K}{\sum_{t=1}^{T-1}K_t} \to 0$ and $\frac{\sum_{t=1}^{T-1}\sqrt{K_t}}{\sqrt{n(T-1)}} \to 0$, $\hat{\theta}^1_{2sl,nT}$ is asymptotically centered normal. ■

## D.4 Proof for Theorem 4

We will first prove the consistency of the GMME under $\frac{\sum_{t=1}^{T-1}K_t}{n(T-1)} \to 0$, then establish its asymptotic normality. Subsequently, we analyze its bias corrected version.

For the identification, Assumption 10 provides the sufficient rank condition. Based on the ideal IVs, $\lambda_0$ and $\delta_0$ can be identified. For the many IVs approach, as linear combinations of the many IVs converge to the ideal IVs in the limit from Lemma 10, $\lambda_0$ and $\delta_0$ can thus be identified from the many IV's conditions. For the uniform convergence of $g'_{nT}(\theta)\Sigma_{nT}^{-1}g_{nT}(\theta)$ in $\theta$, as analysis of the part $g'_{nT,1}(\theta)\Sigma_{nT,1}^{-1}g_{nT,1}(\theta)$ is the same as that in Appendix D.1, we analyze the remaining $g'_{nT,2}(\theta)\Sigma_{nT,2}^{-1}g_{nT,2}(\theta) = \sum_{t=1}^{T-1} V^{*\prime}_{nt}(\theta)M_{nt}V^{*}_{nt}(\theta)$. As $V^*_{nt}(\theta) = d^*_{nt}(\theta) + S_n(\lambda)S_n^{-1}V^*_{nt}$ with $d^*_{nt}(\theta) = (\lambda_0-\lambda)G_nZ^*_{nt}\delta_0 + Z^*_{nt}(\delta_0-\delta)$, we have

$$\frac{1}{n(T-1)}\sum_{t=1}^{T-1} V^{*\prime}_{nt}(\theta)M_{nt}V^{*}_{nt}(\theta) = \frac{1}{n(T-1)}\sum_{t=1}^{T-1} V^{*\prime}_{nt}S_n^{\prime -1}S'_n(\lambda)M_{nt}S_n(\lambda)S_n^{-1}V^*_{nt}$$
$$+\frac{1}{n(T-1)}\sum_{t=1}^{T-1} d^{*\prime}_{nt}(\theta)M_{nt}d^*_{nt}(\theta) + \frac{2}{n(T-1)}\sum_{t=1}^{T-1} d^{*\prime}_{nt}(\theta)M_{nt}S_n(\lambda)S_n^{-1}V^*_{nt}.$$

From Lemma 8, $\frac{1}{n(T-1)}\sum_{t=1}^{T-1} V^{*\prime}_{nt}S_n^{\prime -1}S'_n(\lambda)M_{nt}S_n(\lambda)S_n^{-1}V^*_{nt} \xrightarrow{p} 0$ under $\frac{\sum_{t=1}^{T-1}K_t}{n(T-1)} \to 0$. From Lemma 9, $\frac{1}{n(T-1)}\sum_{t=1}^{T-1}\eta'_{nt}M_{nt}S_n(\lambda)S_n^{-1}V^*_{nt} \xrightarrow{p} 0$ under $\frac{K_T}{n(T-1)} \to 0$; from Lemma 11 (iii), $\frac{1}{n(T-1)}\sum_{t=1}^{T-1} f'_{nt}M_{nt}S_n(\lambda)S_n^{-1}V^*_{nt} = O_p\left(\frac{1}{\sqrt{n(T-1)}}\right)$. Thus, by $(W_nY^*_{nt}, Z^*_{nt}) = f_{nt} + u_{nt}$, $\frac{2}{n(T-1)}\sum_{t=1}^{T-1} d^{*\prime}_{nt}(\theta)M_{nt}S_n(\lambda)S_n^{-1}V^*_{nt} \xrightarrow{p} 0$ under $\frac{\sum_{t=1}^{T-1}K_t}{n(T-1)} \to 0$. Also, as $\frac{1}{n(T-1)}\sum_{t=1}^{T-1} d^{*\prime}_{nt}(\theta)M_{nt}d^*_{nt}(\theta) = (\lambda_0-\lambda,(\delta_0-\delta)')\hat{H}(\lambda_0-\lambda,(\delta_0-\delta)')'$ where $\hat{H}$ has the limit equal to $\text{plim}_{n\to\infty}\mathbf{\Sigma}_{nT,22}$ in Assumption 10, $\frac{1}{n(T-1)}\sum_{t=1}^{T-1} V^{*\prime}_{nt}(\theta)M_{nt}V^*_{nt}(\theta) \xrightarrow{p} \text{plim}_{n\to\infty}\mathbf{\Sigma}_{nT,22}$ uniformly in $\theta$ under $\frac{\sum_{t=1}^{T-1}K_t}{n(T-1)} \to 0$. Therefore, by combining the identification uniqueness and uniform convergence, we obtain the consistency of GMME.

As is derived, the best quadratic moment is to use $\mathbf{P}_{n,T-1}$ in (8). From the Taylor expansion,

$$\sqrt{n(T-1)}(\hat{\theta}_{b,nT}-\theta_0) = -\left[\frac{\partial g'_{nT}(\hat{\theta}_{b,nT})/\partial\theta}{n(T-1)}\Sigma_{nT}^{-1}\frac{\partial g_{nT}(\bar{\theta}_{nT})/\partial\theta'}{n(T-1)}\right]^{-1}\frac{\partial g'_{nT}(\hat{\theta}_{b,nT})/\partial\theta}{n(T-1)}\Sigma_{nT}^{-1}\frac{g_{nT}(\theta_0)}{\sqrt{n(T-1)}}, \quad (39)$$

where $\bar{\theta}_{nT}$ lies between $\hat{\theta}_{b,nT}$ and $\theta_0$. By denoting

$$D_{nT} = -\frac{1}{n(T-1)}\begin{pmatrix} \sigma_0^2 tr(\mathbf{G}'_{n,T-1}\mathbf{P}^s_{n,T-1}) & (\mathbf{W}_{n,T-1}\mathbf{Y}^*_{n,T-1})'\mathbf{H}_{n,T-1} \\ \mathbf{0}_{k_z\times 1} & \mathbf{Z}^{*\prime}_{n,T-1}\mathbf{H}_{n,T-1}\end{pmatrix}', \quad (40)$$

we have $\frac{1}{n(T-1)}\frac{\partial g_{nT}(\hat{\theta}_{nT})}{\partial\theta'} = D_{nT} + o_p(1)$ by Lemma 1 and $\hat{\theta}_{b,nT}-\theta_0 = o_p(1)$. Hence, (39) can be rewritten as $\sqrt{n(T-1)}(\hat{\theta}_{b,nT}-\theta_0) = -\left[D'_{nT}\Sigma_{nT}^{-1}D_{nT}\right]^{-1}D'_{nT}\Sigma_{nT}^{-1}\frac{g_{nT}(\theta_0)}{\sqrt{n(T-1)}} + o_p(1)$. By using $\Sigma_{nT}$ in (21), $D_{nT}$ in (40), and $\frac{g_{nT}(\theta_0)}{\sqrt{n(T-1)}} = \frac{1}{\sqrt{n(T-1)}}\left(\mathbf{V}^{*\prime}_{n,T-1}(\theta)\mathbf{P}_{n,T-1}\mathbf{V}^*_{n,T-1}(\theta), \mathbf{V}^{*\prime}_{n,T-1}(\theta)\mathbf{H}_{n,T-1}\right)'$, we have

$$\begin{aligned} D'_{nT}\Sigma_{nT}^{-1}\frac{g_{nT}(\theta_0)}{\sqrt{n(T-1)}} &= -\frac{1}{\sqrt{n(T-1)}}\begin{pmatrix}\sigma_0^2 tr(\mathbf{G}'_{n,T-1}\mathbf{P}^s_{n,T-1})\left[\boldsymbol{\nu}_P\right]^{-1}\mathbf{V}^{*\prime}_{n,T-1}\mathbf{P}_{n,T-1}\mathbf{V}^*_{n,T-1} \\ \mathbf{0}_{k_z\times 1}\end{pmatrix} \\ &\quad -\frac{1}{\sqrt{n(T-1)}\sigma_0^2}\begin{pmatrix}\sum_{t=1}^{T-1}(W_nY^*_{nt})'M_{nt}V^*_{nt} \\ \sum_{t=1}^{T-1}Z^{*\prime}_{nt}M_{nt}V^*_{nt}\end{pmatrix}, \end{aligned} \quad (41)$$

where $\boldsymbol{\nu}_P = (\mu_4 - 3\sigma_0^4)vec'_D(\mathbf{P}_{n,T-1})vec_D(\mathbf{P}_{n,T-1}) + \sigma_0^4 tr(\mathbf{P}'_{n,T-1}\mathbf{P}^s_{n,T-1})$ and

$$\begin{aligned} D'_{nT}\Sigma_{nT}^{-1}D_{nT} &= \frac{1}{n(T-1)}\begin{pmatrix}\sigma_0^4 tr(\mathbf{G}'_{n,T-1}\mathbf{P}^s_{n,T-1})\left[\boldsymbol{\nu}_P\right]^{-1}tr(\mathbf{G}'_{n,T-1}\mathbf{P}^s_{n,T-1}) & \mathbf{0}_{k_z\times 1} \\ \mathbf{0}_{1\times k_z} & \mathbf{0}_{k_z\times k_z}\end{pmatrix} \\ &\quad +\frac{1}{\sigma_0^2}\frac{1}{n(T-1)}\sum_{t=1}^{T-1}(W_nY^*_{nt}, Z^*_{nt})'M_{nt}(W_nY^*_{nt}, Z^*_{nt}). \end{aligned} \quad (42)$$

The second components of (41) and (42) correspond to $\hat{h}$ and $\hat{H}$ of Appendix D.3. Thus, the analysis in Theorem 3 can be carried over here and we obtain $D'_{nT}\Sigma_{nT}^{-1}\frac{g_{nT}(\theta_0)}{\sqrt{n(T-1)}} \xrightarrow{d} N(0, \text{plim}_{n\to\infty} D'_{nT}\Sigma_{nT}^{-1}D_{nT})$ with $\text{plim}_{n\to\infty} D'_{nT}\Sigma_{nT}^{-1}D_{nT} = \lim_{n\to\infty}\begin{pmatrix} \frac{1}{n(T-1)}tr[\mathbf{P}^s_{n,T-1}\mathbf{G}_{n,T-1}] & \mathbf{0}_{1\times k_z} \\ \mathbf{0}_{k_z\times 1} & \mathbf{0}_{k_z\times k_z}\end{pmatrix} + \frac{1}{\sigma_0^2}\text{plim}_{n\to\infty}\mathbf{\Sigma}_{nT,22} = \mathbf{\Sigma}_b$.

When we use an estimated $\hat{\mathbf{P}}_{n,T-1}$ and $\hat{\Sigma}_{nT}$, the result holds similar to the proof in Appendix D.2. ■

# E Best Quadratic Moment for the Model with Time Dummies

This section derives the best quadratic moment matrix (27) for the model with time dummies. For the covariance of $V^{*'}_{n,T-1}(I_{T-1}\otimes J_nP_nJ_n)V^*_{n,T-1}$ where $tr(P_nJ_n)=0$, from Lemma 2,

$$\begin{aligned} & \text{Cov}(V^{*'}_{n,T-1}(I_{T-1}\otimes J_nP_{n1}J_n)V^*_{n,T-1}\cdot V^{*'}_{n,T-1}(I_{T-1}\otimes J_nP_{n2}J_n)V^*_{n,T-1}) \\ = \ & \sigma_0^4 tr[(I_{T-1}\otimes J_nP^s_{n1}J_n)(I_{T-1}\otimes J_nP_{n2}J_n)] + (\mu_4-3\sigma_0^4)vec'_D(I_{T-1}\otimes J_nP_{n1}J_n)vec_D(I_{T-1}\otimes J_nP_{n2}J_n) \\ = \ & (T-1)\{\sigma_0^4 tr(J_nP^s_{n1}J_nP_{n2}J_n) + (\mu_4-3\sigma_0^4)vec'_D(J_nP_{n1}J_n)vec_D(J_nP_{n2}J_n)\}. \end{aligned}$$

Then, by using Lemmas 12-15 below, the best quadratic matrix, which takes into account $\eta_4$, is $P_n^*$ in (27).

**Lemma 12** *Suppose $tr(P_nJ_n)=0$, then $diag[J_ndiag(J_nP_nJ_n)J_n] = \frac{n-2}{n}diag(J_nP_nJ_n)$.*

**Lemma 13** *Suppose $tr(P_nJ_n)=0$ where $P_n$ is either $P_{n1}$ or $P_{n2}$, then*

$$\begin{aligned} tr(J_nP^s_{n1}J_n\cdot J_nP_{n2}J_n) \ = \ & vec'[J_nP^s_{n1}J_n - J_ndiag(J_nP^s_{n1}J_n)J_n]\cdot vec[J_nP_{n2}J_n - J_ndiag(J_nP_{n2}J_n)J_n] \\ & +2(\frac{n+2}{n})vec'_D(J_nP_{n1}J_n)vec_D(J_nP_{n2}J_n). \end{aligned}$$

**Lemma 14** *There exists a scalar $\alpha$ such that*

$$\begin{aligned} & tr(J_nP^s_{n1}J_n\cdot J_nP_{n2}J_n) + (\eta_4-3)vec'_D(J_nP_{n1}J_n)vec_D(J_nP_{n2}J_n) \\ = \ & \frac{1}{2}\{vec'[J_nP^s_{n1}J_n + (\alpha-1)J_ndiag(J_nP^s_{n1}J_n)J_n]\cdot vec[J_nP^s_{n2}J_n + (\alpha-1)J_ndiag(J_nP^s_{n2}J_n)J_n]\}, \end{aligned}$$

*where $\alpha$ solves the quadratic equation $(\frac{n-2}{n})\alpha^2 + \frac{4}{n}\alpha = \frac{\eta_4-3}{2} + \frac{n+2}{n}$. Explicitly, $\alpha$ can be taken as*

$$\alpha_n = -(\frac{2}{n-2}) + \sqrt{\frac{n}{n-2}}\sqrt{\frac{\eta_4-3}{2} + \frac{n}{n-2}}.$$

**Lemma 15** *(i) There exists a diagonal matrix $A_n$ with $tr(A_n)=0$ such that*

$$tr(J_nP^s_nJ_nG_nJ_n) = tr\{[J_nP^s_nJ_n + (\alpha_n-1)J_ndiag(J_nP^s_nJ_n)J_n]\cdot J_n(G_n - \frac{tr(G_n)}{n-1}J_n + A_n)J_n\},$$

*where $A_n = \frac{n(1-\alpha_n)}{2+(n-2)\alpha_n}[diag(J_nG_nJ_n) - \frac{tr(G_nJ_n)}{n}I_n]$.*

*(ii) Let* $P_n^* = (G_n - \frac{tr(J_nG_n)}{n-1}J_n) + \frac{(1-\alpha_n)^2}{(\frac{n}{n-2}+\frac{\eta_4-3}{2})}[diag(J_nG_nJ_n) - \frac{tr(G_nJ_n)}{n}I_n]$*, which has* $tr(P_n^*J_n) = 0$*. Then,* $J_nP_n^*J_n + (\alpha_n - 1)J_ndiag(J_nP_n^*J_n)J_n = J_n(G_n - \frac{tr(G_n)}{n-1}J_n + A_n)J_n$.

**Proof for Lemma 12:** Let $D_n$ be a diagonal matrix. We have $J_nD_nJ_n = D_n - \frac{1}{n}l_nl_n'D_n - \frac{1}{n}D_nl_nl_n' + \frac{tr(D_n)}{n^2}l_nl_n'$. As $diag(l_nl_n') = I_n$ and $diag(A_nD_n) = diag(A_n)\cdot D_n$, we have $diag[J_ndiag(J_nP_nJ_n)J_n] = (1-\frac{2}{n})diag(J_nP_nJ_n)$ because $tr(J_nP_nJ_n) = tr(P_nJ_n) = 0$. ■

**Proof for Lemma 13:** We have

$$\begin{aligned}
& vec'[J_nP_{n1}^sJ_n - J_ndiag(J_nP_{n1}^sJ_n)J_n]\cdot vec[J_nP_{n2}J_n - J_ndiag(J_nP_{n2}J_n)J_n] \\
= \;& tr\{[J_nP_{n1}^sJ_n - J_ndiag(J_nP_{n1}^sJ_n)J_n]\cdot[J_nP_{n2}J_n - J_ndiag(J_nP_{n2}J_n)J_n]\} \\
= \;& tr(J_nP_{n1}^sJ_nP_{n2}) - (1+\frac{2}{n})tr[diag(J_nP_{n1}^sJ_n)diag(J_nP_{n2}J_n)].
\end{aligned}$$

Therefore,

$$\begin{aligned}
& tr(J_nP_{n1}^sJ_n\cdot J_nP_{n2}J_n) - vec'[J_nP_{n1}^sJ_n - J_ndiag(J_nP_{n1}^sJ_n)J_n]\cdot vec[J_nP_{n2}J_n - J_ndiag(J_nP_{n2}J_n)J_n] \\
= \;& (1+\frac{2}{n})tr[diag(J_nP_{n1}^sJ_n)diag(J_nP_{n2}J_n)] = 2(\frac{n+2}{n})vec_D'(J_nP_{n1}J_n)vec_D(J_nP_{n2}J_n). \;■
\end{aligned}$$

**Proof for Lemma 14:** From Lemma 13,

$$\begin{aligned}
& tr(J_nP_{n1}^sJ_n\cdot J_nP_{n2}J_n) + (\eta_4-3)vec_D'(J_nP_{n1}J_n)vec_D(J_nP_{n2}J_n) \\
= \;& \frac{1}{2}\left\{\begin{array}{c} vec'[J_nP_{n1}^sJ_n - J_ndiag(J_nP_{n1}^sJ_n)J_n]\cdot vec[J_nP_{n2}^2J_n - J_ndiag(J_nP_{n2}^2J_n)J_n] \\ +(\frac{\eta_4-3}{2}+\frac{n+2}{n})vec_D'(J_nP_{n1}^sJ_n)vec_D(J_nP_{n2}^sJ_n) \end{array}\right\}.
\end{aligned}$$

First, by arrangement, we have

$$\begin{aligned}
& tr(J_nP_{n1}^sJ_n\cdot J_nP_{n2}J_n) + (\eta_4-3)vec_D'(J_nP_{n1}J_n)vec_D(J_nP_{n2}J_n) \\
= \;& \frac{1}{2}\left\{\begin{array}{c} vec'[J_nP_{n1}^sJ_n - J_ndiag(J_nP_{n1}^sJ_n)J_n]\cdot vec[J_nP_{n2}^sJ_n - J_ndiag(J_nP_{n2}^sJ_n)J_n] \\ +[\frac{\eta_4-3}{2}+\frac{n+2}{n}]vec_D'(J_nP_{n1}^sJ_n)vec_D(J_nP_{n2}^sJ_n) \end{array}\right\}
\end{aligned}$$

Next, for any $\alpha$, we have

$$\begin{aligned}
& vec'[J_nP_{n1}^sJ_n + (\alpha-1)J_ndiag(J_nP_{n1}^sJ_n)J_n]\cdot vec[J_nP_{n2}^sJ_n + (\alpha-1)J_ndiag(J_nP_{n2}^sJ_n)J_n] \\
= \;& tr\{[J_nP_{n1}^sJ_n - J_ndiag(J_nP_{n1}^sJ_n)J_n]\cdot[J_nP_{n2}^sJ_n - J_ndiag(J_nP_{n2}^sJ_n)J_n]\} \\
& +\frac{4\alpha}{n}tr[diag(J_nP_{n1}^sJ_n)diag(J_nP_{n2}^sJ_n)] + (1-\frac{2}{n})\alpha^2tr[diag(J_nP_{n1}^sJ_n)diag(J_nP_{n2}^sJ_n)].
\end{aligned}$$

By matching the above relevant expressions, one can determine $\alpha_n$ which provides the equality in the proposition. The $\alpha_n$ is one of the roots which solve the quadratic equation. The root with the plus sign is taken (when $\eta_4 = 3$ under normality, the corresponding solution of $\alpha$ shall be one). ■

**Proof for Lemma 15:** As $tr(P_nJ_n)=0$, $tr(J_nP_n^sJ_nG_nJ_n)=tr[J_nP_n^sJ_n(G_n-\frac{tr(G_nJ_n)}{n-1}J_n)J_n]$. Thus,

$$\begin{aligned}&tr\{[J_nP_n^sJ_n+(\alpha_n-1)diag(J_nP_n^sJ_n)J_n]\cdot J_n(G_n-\frac{tr(G_n)}{n-1}J_n+A_n)J_n\}\\ =\ & tr[J_nP_n^sJ_n(G_n-\frac{tr(G_n)}{n-1}J_n)]+tr(J_nP_n^sJ_nA_n)+(\alpha_n-1)tr\{J_ndiag(J_nP_n^sJ_n)J_n(G_n-\frac{tr(G_n)}{n-1}J_n+A_n)\}.\end{aligned}$$

The $A_n$ needs be solved from the relation

$$tr(J_nP_n^sJ_nA_n)+(\alpha_n-1)tr[J_ndiag(J_nP_n^sJ_n)J_nA_n]+(\alpha_n-1)tr[J_ndiag(J_nP_n^sJ_n)J_n(G_n-\frac{tr(G_n)}{n-1})]=0.$$

If $A_n$ were a diagonal matrix with zero trace, then $diag(J_nA_nJ_n)=(1-\frac{2}{n})A_n$. Hence,

$$tr[diag(J_nP_n^sJ_n)A_n]+(\alpha_n-1)(1-\frac{2}{n})tr[diag(J_nP_n^sJ_n)A_n]+(\alpha_n-1)tr[diag(J_nP_n^sJ_n)diag[J_n(G_n-\frac{tr(G_n)}{n-1})J_n]=0.$$

The $A_n$ stated in the proposition is a diagonal matrix with zero trace, which satisfies this relation. This justifies (i). The result in (ii) can be checked algebraically. ■

# References


Ahn, S.C. and P. Schmidt, 1995. Efficient estimation of models for dynamic panel data. *Journal of Econometrics* 68, 5-27.

Alvarez, J. and M. Arellano, 2003. The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71, 1121-1159.

Anderson, T.W. and C. Hsiao, 1981. Estimation of dynamic models with error components. *Journal of the American Statistical Association* 76, 598-606.

Arellano, M. and O. Bond, 1991. Some tests of specification for panel data: monte carlo evidence and an application to employment equations. *Review of Economic Studies* 58, 277-297.

Arellano, M. and O. Bover, 1995. Another look at the instrumental-variable estimation of error-components models. *Journal of Econometrics* 68, 29-51.

Baltagi, B., S.H. Song, B.C. Jung and W. Koh, 2007. Testing for serial correlation, spatial autocorrelation and random effects using panel data. *Journal of Econometrics* 140, 5-51.

Bekker, P.A., 1994. Alternative Approximations to the distributions of instrumental variable estimators. *Econometrica* 62, 657-681.

Blundell, R. and S. Bond, 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87, 115-143.

Bun, M. and M. Carree, 2005. Bias-corrected estimation in dynamic panel data models. *Journal of Business & Economic Statistics* 3, No.2, 200-210(11).

Bun, M.J.G. and J.F. Kiviet, 2006. The effects of dynamic feedbacks and LS and MM estimator accuracy in panel data models. *Journal of Econometrics* 132, 409-444.

Chao, J.C. and N.R. Swanson, 2005. Consistent estimation with a large number of weak instruments. *Econometrica* 73, 1673-1692.

Donald, C.G. and W.K. Newey, 2001. Choosing the number of instruments. *Econometrica* 69, 1161-1191.

Han, C. and P.C.B. Phillips, 2006. GMM with many moment conditions. *Econometrica* 74, 147-192.

Hansen, L. P., 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029-1054.

Horn, R. and C. Johnson, 1985. *Matrix Algebra*. Cambridge University Press.

Hsiao, C., M. H. Pesaran and A. K. Tahmiscioglu, 2002. Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics* 109, 107-150.

Kelejian, H.H. and I.R. Prucha, 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbance. *Journal of Real Estate Finance and Economics* 17:1, 99-121.

Kelejian H.H. and I.R. Prucha, 2001. On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics* 104, 219-257.

Kelejian, H.H., I.R. Prucha, and E. Yuzefovich, 2004. Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: large and small sample results, in J. LeSage and K. Pace (eds). *Advances in Econometrics: Spatial and Spatiotemporal Econometrics*, Elsevier, New York.

Kiviet, J., 1995. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics* 68, 81-126.

Korniotis, G.M., 2008. Estimating panel models with internal and external habit formation. *Journal of Business and Economic Statistics*, forthcoming.

Lee, L.F., 2003. Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews* 22, 307–335.

Lee, L.F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models. *Econometrica* 72, 1899-1925.

Lee, L.F., 2007. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics* 137, 489-514.

Lee, L.F. and X. Liu, 2010. Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. *Econometric Theory* 26, 187-230.

Lee, L.F. and J. Yu, 2007. Near unit root in spatial autoregressive model. Manuscript, Ohio State University.

Lee, L.F. and J. Yu, 2010. A spatial dynamic panel data model with both time and individual fixed effects. *Econometric Theory* 26, 564-597.

Liu, X, L. Lee and C. Bollinger, 2009. Improved efficient quasi maximum likelihood sstimator of spatial autoregressive models. Manuscript, Ohio State University.

Moran, P., 1950. Notes on continuous stochastic phenomena. *Biometrika* 37, 17-23.

Neyman, J. and E. Scott, 1948. Consistent estimates based on partially consistent observations. *Econometrica* 16, No.1, 1-32.

Nickell, S.J., 1981. Biases in dynamic models with fixed effects. *Econometrica* 49, 1417-1426.

Okui, R., 2009. The optimal choice of moments in dynamic panel data models. *Journal of Econometrics* 151, 1-16.

Su, L. and Z. Yang., 2007. QML estimation of dynamic panel data models with spatial errors. Manuscript, Beijing University and Singapore Management University.

Yu, J., R. de Jong and L.F. Lee, 2007. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both $n$ and $T$ are large: a nonstationary case. Manuscript, Ohio State University.

Yu, J., R. de Jong and L.F. Lee, 2008. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both $n$ and $T$ are large. *Journal of Econometrics* 146, 118-134.

Yu, J. and L.F. Lee, 2010. Estimation of unit root spatial dynamic data models. *Econometric Theory*, forthcoming.